BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

OM protein - protein search, using sw model

Run on: May 2, 2003, 09:34:42 ; Search time 285 Seconds
(without alignments)
55.302 Million cell updates/sec

Title: US-10-020-674-2
Perfect score: 2013
Sequence: 1 MKGLLYYGTNDIRYSETVPE.....QLIEHKENNVKILVTPNEVS 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	923.5	45.9	382	1	BDH_YEAST	P39714 saccharomyc
2	912	45.3	417	1	YAG1_YEAST	P39713 saccharomyc
3	408.5	20.3	352	1	DHSO_BACSU	Q06004 bacillus su
4	370.5	18.4	348	1	DHSO_BOMMO	Q02912 bombyx mori
5	362	18.0	360	1	DHSO_SCHPO	P36624 schizosacch
6	360.5	17.9	357	1	DHSO_YEAST	P35497 saccharomyc
7	342	17.0	358	1	YDJL_ECOLI	P77539 escherichia
8	342	17.0	363	1	XYL2_PICST	P22144 pichia stip
9	338	16.8	356	1	DHSO_HUMAN	Q00796 homo sapien
10	337.5	16.8	343	1	DHSO_BACHD	Q9z9u1 bacillus ha
11	337	16.7	347	1	YDJJ_ECOLI	P77280 escherichia
12	336.5	16.7	354	1	DHSO_SHEEP	P07846 ovis aries
13	331.5	16.5	399	1	DHSO_RAT	P27867 rattus norv
14	329	16.3	375	1	DHSO_MOUSE	Q64442 mus musculu
15	319.5	15.9	339	1	RSPB_ECOLI	P38105 escherichia
16	316	15.7	337	1	YJJN_ECOLI	P39400 escherichia
17	310	15.4	344	1	TDH_RHIME	Q52998 rhizobium m
18	307.5	15.3	346	1	GATD_ECOLI	P37190 escherichia
19	306	15.2	342	1	Y053_HAEIN	Q57517 haemophilus
20	306	15.2	424	1	FADH_METMR	P47734 methylobact
21	297	14.8	347	1	TDH_BACSU	O31776 bacillus su
22	279.5	13.9	340	1	TDH_XANCP	O34268 xanthomonas
23	271	13.5	412	1	YBDR_ECOLI	P77316 escherichia
24	269	13.4	341	1	TDH_ECOLI	P07913 escherichia
25	265.5	13.2	343	1	IDND_ECOLI	P39346 escherichia
26	261	13.0	379	1	FADH_PICPA	O74685 pichia past
27	257	12.8	339	1	ADH3_BACST	P42328 bacillus st
28	257	12.8	369	1	ADH3_SYNY3	P73138 synechocyst
29	255.5	12.7	375	1	ADH2_PERMA	P41681 peromyscus
30	255	12.7	369	1	ADH3_ECOLI	P25437 escherichia
31	253.5	12.6	360	1	FADH_AMYME	P80094 amycolatops
32	250	12.4	359	1	MTD2_ARATH	Q02972 arabidopsis
33	248	12.3	337	1	ADH1_BACST	P12311 bacillus st
34	246	12.2	336	1	ADHP_ECOLI	P39451 escherichia
35	246	12.2	337	1	ADH1_ZYMMO	P20368 zymomonas m
36	246	12.2	339	1	ADH2_BACST	P42327 bacillus st
37	245	12.2	357	1	MTD1_ARATH	Q02971 arabidopsis
38	245	12.2	379	1	ADH2_STRCA	P80468 struthio ca
39	244	12.1	378	1	ADH3_HAEIN	P44557 haemophilus
40	243	12.1	376	1	ADHX_SPAAU	P79896 sparus aura
41	242.5	12.0	353	1	YPHC_ECOLI	P77360 escherichia
42	241.5	12.0	386	1	FADH_YEAST	P32771 saccharomyc
43	239.5	11.9	366	1	ADH_ALCEU	P14940 alcaligenes
44	239	11.9	378	1	ADHX_DROME	P46415 drosophila
45	238.5	11.8	350	1	ADH_SCHPO	P00332 schizosacch

ALIGNMENTS

RESULT 1
BDH_YEAST
ID BDH_YEAST STANDARD; PRT; 382 AA.
AC P39714;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE (R,R)-butanediol dehydrogenase (EC 1.1.1.4).
GN BDH OR YAL060W OR FUN49.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=FY834alpha;
RX MEDLINE=20549593; PubMed=10938079;
RA Gonzalez E., Fernandez M.R., Larroy C., Sola L., Pericas M.A.,
RA Pares X., Biosca J.A.;
RT "Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the
RT Saccharomyces cerevisiae YAL060W gene product. Disruption and
RT induction of the gene.";
RL J. Biol. Chem. 275:35876-35885(2000).
CC -!- CATALYTIC ACTIVITY: (R,R)-butane-2,3-diol + NAD(+) = (R)-acetoin +
CC NADH.
CC -!- COFACTOR: Zinc (By similarity).
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC ---

CC ---
DR EMBL; U12980; AAC04974.1; -.
DR SGD; S0000056; BDH.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD.
FT METAL 39 39 ZINC 1 (BY SIMILARITY).

```
FT   METAL        73     73       ZINC 1 (BY SIMILARITY).
FT   METAL       103    103       ZINC 2 (BY SIMILARITY).
FT   METAL       120    120       ZINC 2 (BY SIMILARITY).
FT   METAL       123    123       ZINC 2 (BY SIMILARITY).
FT   METAL       131    131       ZINC 2 (BY SIMILARITY).
FT   METAL       173    173       ZINC 1 (BY SIMILARITY).
SQ   SEQUENCE   382 AA;  41538 MW;  1006FA596DE91A17 CRC64;

  Query Match             45.9%;  Score 923.5;  DB 1;  Length 382;
  Best Local Similarity   48.7%;  Pred. No. 1.1e-63;
  Matches  184;  Conservative   65;  Mismatches  126;  Indels    3;  Gaps    2;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
Db       1 MRALAYFKKGDIHFTNDIPRPEIQTDDEVIIDVSWCGICGSDLHE--YLDGPIFMPKDGE 58

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
Db      59 CHKLSNAALPLAMGHEMSGIVSKVGPKVTKVKVGDHVVVDAASSCADLHCWPHSKFYNSK 118

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
Db     119 PCDACQRGSENLCTHAGFVGLGVISGGFAEQVVVSQHHIIPVPKEIPLDVAALVEPLSVT 178

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
Db     179 WHAVKISGFKKGSSALVLGAGPIGLCTILVLKGMGASKIVVSEIAERRIEMAKKLGVEVF 238

Qy     241 DPSTCDDAN-AVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
Db     239 NPSKHGHKSIEILRGLTKSHDGFDYSYDCSGIQVTFETSLKALTFKGTATNIAVWGPKPV 298

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
Db     299 PFQPMDVTLQEKVMTGSIGYVVEAFEEVVRAIHNGDIAMEDCKQLITGKQRIEDGWEKGF 358

Qy     360 KQLIEHKENNVKILVTPN 377
Db     359 QELMDHKESNVKILLTPN 376


RESULT 2
YAG1_YEAST
ID   YAG1_YEAST     STANDARD;      PRT;   417 AA.
AC   P39713;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical zinc-type alcohol dehydrogenase-like protein in GDH3-CNE1
DE   intergenic region.
GN   YAL061W OR FUN50.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC   -!- COFACTOR: ZINC (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12980; AAC04973.1; -.
DR   SGD; S0000057; YAL061W.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Hypothetical protein; Oxidoreductase; Zinc.
FT   METAL        39     39       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        64     64       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       120    120       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       123    123       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       131    131       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       173    173       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   417 AA;  46098 MW;  D921CEDBE61D0151 CRC64;

  Query Match             45.3%;  Score 912;  DB 1;  Length 417;
  Best Local Similarity   45.9%;  Pred. No. 9.2e-63;
  Matches  174;  Conservative   71;  Mismatches  130;  Indels    4;  Gaps    2;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
Db       1 MRALAYFGKGNIRFTNHLKEPHIVAPDELVIDIEWCGICGTDLHE--YTDGPIFFPEDGH 58

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
Db      59 THEISHNPLPQAMGHEMAGTVLEVGPGVKNLKVGDKVVVEPTGTCRDRYRWPLSPNVDKE 118

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
Db     119 WCAACKKGYYNICSYLGLCGAGVQSGGFAERVVMNESHCYKVPDFVPLDVAALIQPLAVC 178

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
Db     179 WHAIRVCEFKAGSTALIIGAGPIGLGTILALNAAGCKDIVVSEPAKVRRELAEKMGARVY 238

Qy     241 DPS--TCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHP 298
Db     239 DPTAHAAKESIDYLRSIADGGDGFDYTFDCSGLEVTLNAAIQCLTFRGTAVNLAMWGHHK 298

Qy     299 IGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358
Db     299 IQFSPMDITLHERKYTGSMCYTHHDFEAVIEALEEGRIDIDRARHMITGRVNIEDGLDGA 358

Qy     359 FKQLIEHKENNVKILVTPN 377
Db     359 IMKLINEKESTIKIILTPN 377


RESULT 3
DHSO_BACSU
ID   DHSO_BACSU     STANDARD;      PRT;   352 AA.
AC   Q06004;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase)
DE   (Glucitol dehydrogenase).
GN   GUTB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.
RC   STRAIN=168;
RX   MEDLINE=93094198; PubMed=1460002;
RA   Ng K., Ye R., Wu X.-C., Wong S.-L.;
RT   "Sorbitol dehydrogenase from Bacillus subtilis. Purification,
RT   characterization, and gene cloning.";
RL   J. Biol. Chem. 267:24989-24994(1992).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: REDUCES GLUCITOL TO FRUCTOSE.
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96947; AAA22508.1; -.
DR   EMBL; Z99107; CAB12434.1; -.
DR   PIR; A45052; A45052.
DR   HSSP; P07846; 1SDG.
DR   SubtiList; BG10177; gutB.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Complete proteome.
FT   INIT_MET      0      0
FT   METAL        44     44       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        69     69       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        99     99       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       102    102       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       105    105       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       113    113       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       155    155       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   352 AA;  38259 MW;  D68828A1D8EDFAF0 CRC64;

  Query Match             20.3%;  Score 408.5;  DB 1;  Length 352;
  Best Local Similarity   27.9%;  Pred. No. 3.4e-24;

    Matches  109;  Conservative   66;  Mismatches  155;  Indels   61;  Gaps   11;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||  : : | :|:  ||:| |:| | ::| |||   ||||:||  :           |
Db       8 MKAAVMHNTREIKI-ETLPVPDI-NHDEVLIKVMAVGICGSDLHYY----------T 52

Qy      61 KDKISGY--ELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQD 118
             :|  |  | |   ||| :| :  ||| |   | |||||||    |
Db      53 NGRIGNYVVEKPFILGHECAGEIAAVGSSVDQFKVGDRVAVEPGVTC---------- 99

Qy     119 LGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPIS 178
           | | ||: |  | |  : |       | | :|:   :| :  :|||:  :  ||:|| |
Db     100 -GRCEACKEGRYNLCPDVQFLATPPVDGAFVQYIKMRQDFVFLIPDSLSYEEAALIEPFS 158

Qy     179 VAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE 238
           |  ||  | : ||| |  ::| ||:||  : | :   || |: ::   :| : .||::||
Db     159 VGIHAAARTKLQPGSTIAIMGMGPVGLMAVAAAKAFGAGTIIVTDLEPLRLEAAKKMGAT 218

Qy     239 VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSG---------- 286
                 ||   :|  :  : |   |:: :| |    ::: :    |
Db     219 HIINIREQDALEEIKT-ITNDRGVDVAWETAGNPAALQSALASVRRGGKLAIVGLPSQNE 277

Qy     287 IAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMIT 346
           | :||    |: |    :   |  | |           :: |  |::     : ::|
Db     278 IPLNVPFIADNEIDIYGI-FRYANTYPKG----------IEFLASGIVD---TKHLVT 321

Qy     347 GKVHLKDGVEKGFKQLIEHKENNVKILVTPN 377
             :  |:   :   :: :: |   :|::| ||
Db     322 DQYSLEQ-TQDAMERALQFKNECLKVMVYPN 351


RESULT 4
DHSO_BOMMO
ID   DHSO_BOMMO     STANDARD;      PRT;   348 AA.
AC   Q02912;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase).
GN   SDH.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC   Ditrysia; Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93279312; PubMed=8504807;
RA   Niimi T., Yamashita O., Yaginuma T.;
RT   "A cold-inducible Bombyx gene encoding a protein similar to mammalian
RT   sorbitol dehydrogenase. Yolk nuclei-dependent gene expression in
RT   diapause eggs.";
RL   Eur. J. Biochem. 213:1125-1131(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokai X Asahi; TISSUE=Fat body;
RX   MEDLINE=97087160; PubMed=8933178;
RA   Niimi T., Yamashita O., Yaginuma T.;
RT   "Structure of the Bombyx sorbitol dehydrogenase gene: a possible
RT   alternative use of the promoter.";
RL   Insect Mol. Biol. 5:269-280(1996).
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- INDUCTION: BY COLD.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D13371; BAA02634.1; -.
DR   EMBL; D66906; BAA11030.1; -.
DR   PIR; S32484; S32484.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD.
FT   METAL        40     40       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        65     65       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       151    151       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   348 AA;  37158 MW;  2985304E9ADE6FF0 CRC64;

  Query Match             18.4%;  Score 370.5;  DB 1;  Length 348;
  Best Local Similarity   30.7%;  Pred. No. 2.8e-21;
  Matches  103;  Conservative   56;  Mismatches  146;  Indels   31;  Gaps   10;

Qy       7 YGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISG 66
           :| ||:|  | :| ||| | ::| ||:   ||||:|:|   || |          | |
Db      10 HGANDVRI-EKIPVPEI-NDDEVLIKIDCVGICGSDVK--LYSTGTC------GADVI-- 57

Qy      67 YELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQ 126
           : |:  ||| :||||:||  |:|:|:: ||||||:| |  |  ||           |  |:
Db      58 -DKPIVIGHEGAGTVVKVGDKVSSLRVGDRVAIEPTQPC--RS-----------CELCK 102

Qy     127 SGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVER 186
           |  | |    :|   || |    |  :  |   ||||::  : || |:|:::  ||  |
Db     103 RGKYNLCVEPRYCSSMGAPGNLCRYYKHVADFCHKLPDNLTMEEGAAVQPLAIVIHACNR 162

Qy     187 ARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE--VFDPST 244
           |:   |   ::|| ||||:   :: :   | ||: ::    |   | ||||:  :
Db     163 AKITLGSKIVILGAGPIGILCAMSAKAMGASKIILTDVVQSRLDAALELGADNVLLVRRE 222

Qy     245 CDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPM 304
             |  | | :    :   : | |       ::: |  :|: : | :  |  :  :|:
Db     223 YTDEEVVEKIVKLLGDRPDVSIDACGYGSAQRVALLVTKTAGLVLVVGI-ADKTVE-LPL 280

Qy     305 SLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDK 340
           |    :        :  :|  : |:  | | |||
Db     281 SQALLREVDVVGSFRIMNTYQPALAAVSSGAIPLDK 316


RESULT 5
DHSO_SCHPO
ID   DHSO_SCHPO     STANDARD;      PRT;   360 AA.
AC   P36624;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol
DE   2-dehydrogenase) (Protein tms1).
GN   TMS1 OR SPC1773.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE OF 5-360 FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=94039112; PubMed=8223615;
RA   Wagner P., Grimaldi M., Jenkins J.R.;
RT   "Putative dehydrogenase tms1 suppresses growth arrest induced by a
RT   p53 tumour mutant in fission yeast.";
RL   Eur. J. Biochem. 217:731-736(1993).
CC   -!- FUNCTION: SUPPRESSES GROWTH ARREST INDUCED BY A P53 TUMOR MUTANT
CC       IN FISSION YEAST.
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL033389; CAA21910.1; -.
DR   EMBL; X74422; CAA52443.1; ALT_INIT.
DR   PIR; S35981; S35981.
DR   PIR; S38345; S38345.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD.
FT   METAL        42     42       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        67     67       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       153    153       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   360 AA;  38851 MW;  ADEB3DDBA163224C CRC64;

  Query Match             18.0%;  Score 362;  DB 1;  Length 360;
  Best Local Similarity   29.7%;  Pred. No. 1.3e-20;
  Matches  109;  Conservative   59;  Mismatches  161;  Indels   38;  Gaps    7;

Qy      16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGH 75
           |  |   : : : ||: :   ||||:|:  :   |   |  |:           |: ||
Db      19 EDRPGQTLTDDHQVKVAIKATGICGSDVHYWKEGGIGDFILKK----------PMILGH 67

Qy      76 EFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCAS 135
           | :| ||||| ||:|:|||| ||||    |              ||  |:||  | |
Db      68 ESAGVVVEVGKGVSSLKPGDPVAVEPGCVCR-------------LCDYCRSGRYNLCPH 113

Qy     136 LSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTA 195
           : |       |    | :  ||   |||  |  : ||| ||:||| ||: |   : |
```

```
Db    114 MEFAATPPYDGTLRTYYITTEDFCTKLPKQISVEEGALFEPMSVAVHAMTRGNLKCGSRV 173

Qy    196 LVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFA-KELGAEVFDP-------STCDD 247
          ||:| | :|| :   : : | || : :  | :|| | :||: | |       | |
Db    174 LVMGCGTVGLLMMAVAKAYGAIDIVAVDASPSRVEFAQKYVGAKPFTPIAAKENESLPDY 233

Qy    248 ANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLT 307
          |    :|:: :   |  | | :||     |:::|    | ||  |  |   | |    :
Db    234 AQRYKQAIIEKYGEFDFAVDATGVGICIHTAVLALKRGGTFVQ-AGNGKPVIDFPINHII 292

Qy    308 YQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKE 367
            |    ||  |   ::: :  : :||: :   : :|| :   || : | :: :   :|
Db    293 NYEINVLGSFRYAHGCYKQSLFLVSNGLVDV---KPLITHRFAFKDAL-KAYETVASGEE 348

Qy    368 NNVKILV 374
            :|:::
Db    349 GVLKVII 355


RESULT 6
DHSO_YEAST
ID   DHSO_YEAST     STANDARD;      PRT;   357 AA.
AC   P35497;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase).
GN   SOR1 OR SDH1 OR YJR159W OR J2395.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94171068; PubMed=8125328;
RA   Sarthy A.V., Schopp C., Idler K.B.;
RT   "Cloning and sequence determination of the gene encoding sorbitol
RT   dehydrogenase from Saccharomyces cerevisiae.";
RL   Gene 140:121-126(1994).
RN   [2]
RP   SEQUENCE OF 1-20 FROM N.A.
RA   Huang M.-E., Chuat J.-C., Galibert F.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 11-357 FROM N.A.
RA   Wedler H., Underwood A.P., Louis E.J., Wambutt R.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11039; AAA35027.1; -.
DR   EMBL; Z49659; CAA89692.1; -.
DR   HSSP; P07846; 1SDG.
DR   SGD; S0003920; SOR1.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD.
FT   METAL        43     43       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        68     68       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       154    154       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   357 AA;  38165 MW;  16843C2040A3334A CRC64;

  Query Match            17.9%;  Score 360.5;  DB 1;  Length 357;
  Best Local Similarity   30.3%;  Pred. No. 1.7e-20;
  Matches  111;  Conservative   56;  Mismatches  160;  Indels   39;  Gaps   11;

Qy     16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYEL--PLCP 73
          | | | ||:|: ||: :   ||||:|:  :  |||            | | |  |:
Db     20 EQRPIPTIKDPHYVKLAIKATGICGSDI-HYYRSGG-----------IGKYILKAPMVL 66

Qy     74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
          ||| || |||||  || || |||||:|       ||| |           : |  | |
Db     67 GHESSGQVVEVGDAVTRVKVGDRVAIEP----GVPSRYSDET----------KEGRYNLC 112

Qy    134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
            ::|       |   :| :  || :||||: :  : || |||:||  |: : |  : |
Db    113 PHMAFAATPPIDGTLVKYYLSPEDFLVKLPEGVSYEEGACVEPLSVGVHSNKLAGVRFGT 172

Qy    194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPS--TCDDANA 250
            :| | ||:|| |    :   |  ::  :    : | ||: ||   |: |  : | |
Db    173 KVVVFGAGPVGLLTGAVARAFGATDVIFVDVFDNKLQRAKDFGATNTFNSSQFSTDKAQD 232

Qy    251 VLKAMVPENEGFHA--AFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTY 308
          :  :     | ||   |:|||       ::  |   |  | | : | :   |    ::
Db    233 LADGVQKLLGGNHADVVFECSGADVCIDAAVKTTKVGGTMVQVGM-GKNYTNFPIAEVSG 291

Qy    309 QEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKEN 368
          :|    |   |: |:|: |  :  | :::   : :|| |   :|   | :   | |
Db    292 KEMKLIGCFRYSFGDYRDAVNLVATGKVNV---KPLITHKFKFEDAA-KAYDYNIAHGGE 347

Qy    369 NVKILV 374
            || ::
Db    348 VVKTII 353


RESULT 7
YDJL_ECOLI
ID   YDJL_ECOLI     STANDARD;      PRT;   358 AA.
AC   P77539;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical zinc-type alcohol dehydrogenase-like protein ydjL.
GN   YDJL OR B1776.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- COFACTOR: ZINC (Potential).
```

```
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000272; AAC74846.1; -.
DR   EMBL; D90821; BAA15574.1; -.
DR   EcoGene; EG13488; ydjL.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Hypothetical protein; Oxidoreductase; Zinc; Complete proteome.
FT   METAL        39     39       ZINC (CATALYTIC).
FT   METAL        61     61       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        92     92       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        95     95       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        98     98       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       106    106       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       157    157       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   358 AA;  38732 MW;  B45C8C35B20ED24D CRC64;

  Query Match             17.0%;  Score 342;  DB 1;  Length 358;
  Best Local Similarity   28.2%;  Pred. No. 4.4e-19;
  Matches  110;  Conservative   66;  Mismatches  162;  Indels   52;  Gaps   16;

Qy       1 MKGLLYYGTNDIRYSE-TVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQG 59
           || |  :|     |    ||:| :  | || |::    ||| |:| :    |
Db       1 MKALARFGKAFGGYKMIDVPQP-MCGPEDVVIEIKAAAICGADMKHYNVDSG-------- 51

Qy      60 TKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDL 119
             |: :       ||||:| : :||  |   | | ||  : : |
Db      52 -SDEFNSIR-----GHEFAGCIAQVGEKVKDWKVGQRVVSDNSGHV----------C 92

Qy     120 GLCMACQSGSPNCCASLSFCGLGGAS--GGFAEY-VVYGE------DHMVKLPDSIPDDI 170
           |:| ||: |   ||      ||   :  |||::| :| ||        : ::|| :  :
Db      93 GVCPACEQGDFLCCTEKVNLGLDNNTWGGGFSKYCLVPGEILKIHRHALWEIPDGVDYED 152

Qy     171 GALVEPISVAWHAV-ERARFQPGQTALVLGGGPIGLATILALQGHHAGKIV---CSEPAL 226
           |:::||  |: :: ::::| |||  :|:| ||:|| ::   :   |  ||     |
Db     153 AAVLDPICNAYKSIAQQSKFLPGQDVVVIGTGPLGLFSVQMARIMGAVNIVVVGLQEDVA 212

Qy     227 IRRQFAKELGA-EVFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPS 285
           :| .  ||||||  | : || |      :   :| |     :|||     :|    |:
Db     213 VRFPVAKELGATAVVNGSTEDVVARCQQICGKDNLGL--VIECSGANIALKQAIEMLRPN 270

Qy     286 GIAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMI 345
           |  | | : |  |: |    :|   |  | |    ::  :: | | |   | : ||
Db     271 GEVVRVGM-GFKPLDFSINDITAWNKSIIGHMAYDSTSWRNAIRLLASGAI---KVKPMI 326

Qy     346 TGKVHLKDGVEKGFKQLIEHKENNVKILVT 375
           | :: |    | ||  :::  :  :|:::|
Db     327 THRIGLSQWRE-GFDAMVD--KTAIKVIMT 353


RESULT 8
XYL2_PICST
ID   XYL2_PICST     STANDARD;      PRT;   363 AA.
AC   P22144;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-xylulose reductase (EC 1.1.1.9) (Xylitol dehydrogenase) (XDH).
GN   XYL2.
OS   Pichia stipitis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4924;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 5774;
RX   MEDLINE=91168296; PubMed=2127555;
RA   Koetter P., Amore R., Hollenberg C.P., Ciriacy M.;
RT   "Isolation and characterization of the Pichia stipitis xylitol
RT   dehydrogenase gene, XYL2, and construction of a xylose-utilizing
RT   Saccharomyces cerevisiae transformant.";
RL   Curr. Genet. 18:493-500(1990).
RN   [2]
RP   SIMILARITY TO OTHER ZINC-ALCOHOL DEHYDROGENASES.
RX   MEDLINE=93279386; PubMed=8504864;
RA   Persson B., Hallborn J., Walfridsson M., Hahn-Haegerdal B.,
RA   Keraenen S., Penttilae M., Joernvall H.;
RT   "Dual relationships of xylitol and alcohol dehydrogenases in families
RT   of two protein types.";
RL   FEBS Lett. 324:9-14(1993).
CC   -!- CATALYTIC ACTIVITY: Xylitol + NAD(+) = D-xylulose + NADH.
CC   -!- COFACTOR: ZINC (Potential).
CC   -!- PATHWAY: D-xylose degradation.
CC   -!- INDUCTION: BY XYLOSE. REPRESSED BY GLUCOSE.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55392; CAA39066.1; -.
DR   EMBL; A16166; CAA01269.1; -.
DR   PIR; S13529; S13529.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Xylose metabolism; NAD; Zinc.
FT   METAL        41     41       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        66     66       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       159    159       ZINC (CATALYTIC) (BY SIMILARITY).
FT   NP_BIND     183    188       NAD (POTENTIAL).
SQ   SEQUENCE   363 AA;  38521 MW;  39E16FD087160248 CRC64;

  Query Match             17.0%;  Score 342;  DB 1;  Length 363;
  Best Local Similarity   28.4%;  Pred. No. 4.5e-19;
  Matches  110;  Conservative   56;  Mismatches  159;  Indels   62;  Gaps   14;

Qy      10 NDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYEL 69
           :|| : ||   |||  | || ::|   ||||:|:  : :                :|  : |
Db      13 DDISF-ETYDAPEISEPTDVLVQVKKTGICGSDIHFYAHG------------RIGNFVL 58

Qy      70 --PLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQS 127
             |:  ||| :||||:|| ||||:| || ||:|        ||: |            :|
Db      59 TKPMVLGHESAGTVVQVGKGVTSLKVGDNVAIEP----GIPSRFSDEY----------KS 104

Qy     128 GSPNCCASLSFC-------GLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           |  | |  ::|         |     |   :|    || :||||| :  ::||||||:||
Db     105 GHYNLCPHMAFAATPNSKEGEPNPPGTLCKYFKSPEDFLVKLPDHVSLELGALVEPLSVG 164

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA--E 238
           || :      |   | | ||:||     :   | ::  :    : : ||::||
Db     165 VHASKLGSVAFGDYVAVFGAGPVGLLAAAVAKTFGAKGVIVVDIFDNKLKMAKDIGAATH 224

Qy     239 VFDPSTCDDANAVLKAM---VPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWG 295
           |:  |   :  ::||    ||     :   :|:|        : |  | |  | |
Db     225 TFNSKT-GGSEELIKAFGGNVP-----NVVLECTGAEPCIKLGVDAIAPGGRFVQVGN-A 277
```

```
Qy    296 DHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVV--------KALEDGLISLDKARKMITG 347
          |: |       :|    || |   |:: |           |: |  :  ::||
Db    278 AGPVSFPITVFAMKELTLFGSFRYGFNDYKTAVGIFDTNYQNGRENAPIDFE---QLITH 334

Qy    348 KVHLKDGVEKGFKQLIEHKENNVKILV 374
          :  || :|     |:   :  || |:
Db    335 RYKFKDAIEA--YDLVRAGKGAVKCLI 359


RESULT 9
DHSO_HUMAN
ID   DHSO_HUMAN     STANDARD;      PRT;   356 AA.
AC   Q00796; Q16682; Q9UMD6;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase).
GN   SORD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=90107985; PubMed=2691249;
RA   Karlsson C., Maret W., Auld D.S., Hoeoeg J.-O., Joernvall H.;
RT   "Variability within mammalian sorbitol dehydrogenases. The primary
RT   structure of the human liver enzyme.";
RL   Eur. J. Biochem. 186:543-550(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95301293; PubMed=7782086;
RA   Iwata T., Popescu N.C., Zimonjic D.B., Karlsson C., Hoeoeg J.-O.,
RA   Vaca G., Rodriguez I.R., Carper D.;
RT   "Structural organization of the human sorbitol dehydrogenase gene
RT   (SORD).";
RL   Genomics 26:55-62(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94375058; PubMed=8088829;
RA   Lee F.K., Cheung M.C., Chung S.;
RT   "The human sorbitol dehydrogenase gene: cDNA cloning, sequence
RT   determination, and mapping by fluorescence in situ hybridization.";
RL   Genomics 21:354-358(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97326004; PubMed=9183016;
RA   Carr I.M., Markham A.F., Coletta P.L.;
RT   "Identification and characterisation of a sequence related to human
RT   sorbitol dehydrogenase.";
RL   Eur. J. Biochem. 245:760-767(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29008; AAA80565.1; -.
DR   EMBL; L29254; AAA80566.1; -.
DR   EMBL; L29249; AAA80566.1; JOINED.
DR   EMBL; L29250; AAA80566.1; JOINED.
DR   EMBL; L29251; AAA80566.1; JOINED.
DR   EMBL; L29252; AAA80566.1; JOINED.
DR   EMBL; L29253; AAA80566.1; JOINED.
DR   EMBL; U07361; AAA66064.1; -.
DR   EMBL; U67243; AAB61898.1; -.
DR   EMBL; U67236; AAB61898.1; JOINED.
DR   EMBL; U67237; AAB61898.1; JOINED.
DR   EMBL; U67238; AAB61898.1; JOINED.
DR   EMBL; U67239; AAB61898.1; JOINED.
DR   EMBL; U67240; AAB61898.1; JOINED.
DR   EMBL; U67241; AAB61898.1; JOINED.
DR   EMBL; U67242; AAB61898.1; JOINED.
DR   EMBL; BC021085; AAH21085.1; -.
DR   HSSP; P07846; 1SDG.
DR   Genew; HGNC:11184; SORD.
DR   MIM; 182500; -.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Acetylation; Polymorphism.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   METAL        44     44       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        69     69       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       155    155       ZINC (CATALYTIC) (BY SIMILARITY).
FT   VARIANT     238    238       L -> Q.
FT                                /FTId=VAR_000430.
FT   CONFLICT      4      4       MISSING (IN REF. 1).
FT   CONFLICT     58     58       N -> D (IN REF. 1).
FT   CONFLICT    185    185       M -> E (IN REF. 1).
FT   CONFLICT    280    280       T -> S (IN REF. 1).
FT   CONFLICT    288    288       I -> T (IN REF. 1).
SQ   SEQUENCE   356 AA;  38165 MW;  6BEB067F920A3E09 CRC64;

  Query Match             16.8%;  Score 338;  DB 1;  Length 356;
  Best Local Similarity   30.1%;  Pred. No. 8.9e-19;
  Matches   89;  Conservative   46;  Mismatches  123;  Indels   38;  Gaps    7;

Qy      4 LLYYGTNDIRYSE-TVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKD 62
          |: :|  |:|     :|||    ||:| :::   ||||:|:  : |     |  |:
Db     11 LVVHGPGDLRLENYPIPEP---GPNEVLLRMHSVGICGSDVHYWEYGRIGNFIVKK---- 63

Qy     63 KISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLC 122
                |:  ||| |||| :||| |  :|||||||:|  :      |  |
Db     64 -------PMVLGHEASGTVEKVGSSVKHLKPGDRVAIEPGA-----PRENDEF------- 104

Qy    123 MACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWH 182
           |: |  |   |: ||      |    :  :      ||||::  : |||:||:||  |
Db    105 --CKMGRYNLSPSIFFCATPPDDGNLCRFYKHNAAFCYKLPDNVTFEEGALIEPLSVGIH 162

Qy    183 AVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDP 242
          |  |     |   || | ||||: |:|  :   | ::| :: :  |   |||:||::
Db    163 ACRRGGVTLGHKVLVCGAGPIGMVTLLVAKAMGAAQVVVTDLSATRLSKAKEIGADLVLQ 222

Qy    243 STCDDANAVLKAMVPENEGF-----HAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
           : :    : : :    ||          :|:|   :    | ||   |  | | :
Db    223 ISKESPQEIARKV----EGLLGCKPEVTIECTGAEASIQAGIYATRSGGTLVLVGL 274


RESULT 10
DHSO_BACHD
ID   DHSO_BACHD     STANDARD;      PRT;   343 AA.
AC   Q9Z9U1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase)
DE   (Glucitol dehydrogenase).
```

```
GN   GUTB OR BH0189.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=99184646; PubMed=10086842;
RA   Takami H., Nakasone K., Ogasawara N., Hirama C., Nakamura Y.,
RA   Masui N., Fuji F., Takaki Y., Inoue A., Horikoshi K.;
RT   "Sequencing of three lambda clones from the genome of alkaliphilic
RT   Bacillus sp. strain C-125.";
RL   Extremophiles 3:29-34(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
CC   -!- FUNCTION: REDUCES GLUCITOL TO FRUCTOSE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB011837; BAA75341.1; -.
DR   EMBL; AP001507; BAB03908.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Complete proteome.
FT   METAL        39     39       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        60     60       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        91     91       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        94     94       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        97     97       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       105    105       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       146    146       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   343 AA;  36889 MW;  20961F7C34676DFA CRC64;

  Query Match             16.8%;  Score 337.5;  DB 1;  Length 343;
  Best Local Similarity   28.7%;  Pred. No. 9.2e-19;
  Matches  111;  Conservative   57;  Mismatches  162;  Indels   57;  Gaps   19;

Qy       1 MKGLL--YYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQ 58
           || |:   :||      |  |||     : ||||| | |:||:|:   || |    :|
Db       1 MKALVKTQHGTGHFAVQEK-PEP-TPGKHQVKIKVKYTGVCGSDI--HTYEG---HYP-- 51

Qy      59 GTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEAT-SHCSDRSRYKDTVAQ 117
                  |:  |||||| :||:| |||    |||| | | | |
Db      52 --------VAAPVTLGHEFSGEIVELGEGVTGFNVGDRVTSETTYSIC------------ 91

Qy     118 DLGLCMACQSGSPNCCASLSFCGLGG-ASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEP 176
             | |  | ||  | |:     |||    | ||:||:  :: :  ||  : |   |: ||
Db      92 --GKCSYCTSGDYNLCSHRK--GLGNQQDGSFAKYVIARQESLHHLPAGVDDRSAAMTEP 147

Qy     177 ISVAWHAVERARFQPGQTALVLGGGPIGLATILALQ--GHHAGKIVCS--EPALIRRQFA 232
           ::   ||: :     |   :| | |||||   || |    | | :: :      :| : |

Db     148 LACTHHAIAKTSINKGDLVVVTGPGPIGL---LAAQVAKSHGGTVIITGLSNDQVRLKKA 204

Qy     233 KELGAEVFDPSTCDDANAVLKAMVPE-NEGFHA--AFDCSGVPQTFTTSIVATGPSGIAV 289
           ||:| :       |     :| :| |  :|: |    :|||        |       |
Db     205 KEVGIDY----AIDTQEVDIKELVSELTDGYGADVVLECSGAVPAAKQGIDLLRKKGQYA 260

Qy     290 NVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKV 349
             | ::    | |    :  :|    ||      |::  :  | :  ::   |: ::| :
Db     261 QVGLFAQPEIQFNFEKIIQKEISVVGSRSQKPADWEPALSLLNEKKVN---AKTLVTHEY 317

Qy     350 HLKDGVEKGFKQLIEHKENNVKILVTP 376
           : :  :| :  :       :|:|:||
Db     318 TISEW-DKAYHAI--KSGEAIKVLLTP 341


RESULT 11
YDJJ_ECOLI
ID   YDJJ_ECOLI     STANDARD;      PRT;   347 AA.
AC   P77280; P78174;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical zinc-type alcohol dehydrogenase-like protein ydjJ.
GN   YDJJ OR B1774.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- COFACTOR: ZINC (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000272; AAC74844.1; -.
DR   EMBL; D90820; BAA15565.1; -.
DR   EMBL; D90821; BAA15572.1; -.
DR   HSSP; P07846; 1SDG.
DR   EcoGene; EG13486; ydjJ.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
```

```
KW   Hypothetical protein; Oxidoreductase; Zinc; Complete proteome.
FT   METAL        39     39       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        65     65       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        95     95       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        98     98       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       101    101       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       109    109       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       152    152       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   347 AA;  37701 MW;  F8E2172184626A3B CRC64;

  Query Match             16.7%;  Score 337;  DB 1;  Length 347;
  Best Local Similarity   30.6%;  Pred. No. 1e-18;
  Matches  113;  Conservative   51;  Mismatches  149;  Indels   56;  Gaps   14;

Qy       8 GTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGY 67
Db      12 GTMKI-ISAEIPVPK---EDEVLIKVEYVGICGSDVHGF--ESGPFIPPKDPNQ------ 59

Qy      68 ELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQS 127
Db      60 EIGL--GHECAGTVVAVGSRVRKFKPGDRVNIEPGVPC----------GHCRYCLE 103

Qy     128 GSPNCCASLSFCGL-GGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVER 186
Db     104 GKYNICPDVDFMATQPNYRGALTHYLCHPESFTYKLPDNMDTMEGALVEPAAVGMHAAML 163

Qy     187 ARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDPSTCD 246
Db     164 ADVKPGKKIIILGAGCIGLMTLQACKCLGATEIAVVDVLEKRLAMAEQLGATVVINGAKE 223

Qy     247 DANAVLKAMVPENEGFHAAFDCSGVPQTFTTS--IVATGPSGIAVNVAVWGDHPIGFMPM 304
Db     224 DTIARCQQFT-EDMGADIVFETAGSAVTVKQAPYLVMRG-GKIMIVGTVPGDSAINFLKI 281

Qy     305 S--------LTYQEKY-------ATG-----SMCYTVKDFQEVVKALEDGLISLDKARKM 344
Db     282 NREVTIQTVFRYANRYPVTIEAISSGRFDVKSMVTHIYDYRDVQQAFEE---SVNNKRDI 338

Qy     345 ITGKVHLKD 353
Db     339 IKGVIKISD 347


RESULT 12
DHSO_SHEEP
ID   DHSO_SHEEP     STANDARD;      PRT;   354 AA.
AC   P07846;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase).
GN   SORD.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=84158607; PubMed=6705798;
RA   Jeffery J., Cederlund E., Joernvall H.;
RT   "Sorbitol dehydrogenase. The primary structure of the sheep-liver
RT   enzyme.";
RL   Eur. J. Biochem. 140:7-16(1984).
RN   [2]
RP   SIMILARITY TO ZINC ALCOHOL DEHYDROGENASES.
RX   MEDLINE=84158592; PubMed=6368230;
RA   Joernvall H., von Bahr-Lindstrom H., Jeffery J.;
RT   "Extensive variations and basic features in the alcohol
RT   dehydrogenase-sorbitol dehydrogenase family.";
RL   Eur. J. Biochem. 140:17-23(1984).
RN   [3]
RP   3D-STRUCTURE MODELING.
RX   MEDLINE=86131624; PubMed=2936393;
RA   Eklund H., Horjales E., Jornvall H., Branden C.I., Jeffery J.;
RT   "Molecular aspects of functional differences between alcohol and
RT   sorbitol dehydrogenases.";
RL   Biochemistry 24:8005-8012(1985).
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   PIR; S10065; S10065.
DR   PDB; 1SDG; 15-OCT-94.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Acetylation; 3D-structure.
FT   MOD_RES      1      1       ACETYLATION.
FT   METAL       43     43       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       67     67       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      153    153       ZINC (CATALYTIC) (BY SIMILARITY).
FT   STRAND       4     11
FT   TURN        14     15
FT   STRAND      19     25
FT   TURN        30     31
FT   STRAND      32     41
FT   HELIX       44     51
FT   TURN        52     52
FT   TURN        57     58
FT   STRAND      63     64
FT   STRAND      69     76
FT   TURN        78     79
FT   TURN        85     86
FT   STRAND      88     91
FT   HELIX      101    103
FT   TURN       104    104
FT   TURN       112    113
FT   TURN       116    117
FT   STRAND     127    127
FT   STRAND     131    133
FT   HELIX      134    136
FT   STRAND     137    139
FT   TURN       146    148
FT   HELIX      149    152
FT   TURN       153    153
FT   HELIX      154    163
FT   TURN       164    166
FT   TURN       169    170
FT   STRAND     172    176
FT   HELIX      180    191
FT   STRAND     196    200
FT   HELIX      204    212
FT   TURN       213    214
FT   STRAND     217    219
FT   HELIX      221    223
FT   HELIX      228    235
FT   TURN       236    238
FT   STRAND     242    245
FT   HELIX      250    258
FT   TURN       259    259
FT   TURN       262    264
FT   STRAND     266    269
FT   TURN       282    283
FT   STRAND     287    289
FT   TURN       292    293
FT   HELIX      297    309
FT   TURN       310    311
FT   HELIX      316    318
FT   STRAND     319    321
FT   STRAND     324    324
FT   HELIX      325    335
```

```
FT   STRAND      342    346
SQ   SEQUENCE   354 AA;  37831 MW;  D5FE56BC06F87389 CRC64;

  Query Match           16.7%;  Score 336.5;  DB 1;  Length 354;
  Best Local Similarity   32.4%;  Pred. No. 1.1e-18;
  Matches   97;  Conservative   44;  Mismatches  113;  Indels   45;  Gaps   10;

Qy       4 LLYYGTNDIRYSE-TVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPV--FFPKQGT 60
           |: :|  |:|     :|||    ||:| :|:   ||||:|:    |  | :  |  |:
Db      10 LVVHGPGDLRLENYPIPEP---GPNEVLLKMHSVGICGSDVH---YWQGRIGDFVVKK-- 61

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
                 |:  ||| |||||:||| |  ::|||||::  :      |  |
Db      62 ---------PMVLGHEASGTVVKVGSLVRHLQPGDRVAIQPGA-----PRQTDEF----- 102

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
               |: |  |   :: ||      |    :  :  :   ||||::  : |||:||:||
Db     103 ----CKIGRYNLSPTIFFCATPPDDGNLCRFYKHNANFCYKLPDNVTFEEGALIEPLSVG 158

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA--- 237
            ||  ||    |   || | |||||  :|| :   | ::| :: :  |   |||:||
Db     159 IHACRRAGVTLGNKVLVCGAGPIGLVNLLAAKAMGAAQVVVTDLSASRLSKAKEVGADFI 218

Qy     238 -EVFDPSTCDDANAV--LKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
            |: : |  : |  |  |   ||        :|:||  :    | ||   |  | | :
Db     219 LEISNESPEEIAKKVEGLLGSKPE-----VTIECTGVETSIQAGIYATHSGGTLVLVGL 272


RESULT 13
DHSO_RAT
ID   DHSO_RAT       STANDARD;      PRT;   399 AA.
AC   P27867;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase).
GN   SORD OR SDH1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RC   TISSUE=Liver;
RX   MEDLINE=91266961; PubMed=2050152;
RA   Karlsson C., Joernvall H., Heoeog J.O.;
RT   "Sorbitol dehydrogenase: cDNA coding for the rat enzyme. Variations
RT   within the alcohol dehydrogenase family independent of quaternary
RT   structure and metal content.";
RL   Eur. J. Biochem. 198:761-765(1991).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RC   STRAIN=Sprague-Dawley; TISSUE=Testis;
RX   MEDLINE=94039079; PubMed=8223590;
RA   Wen Y., Bekhor I.;
RT   "Sorbitol dehydrogenase. Full-length cDNA sequencing reveals a mRNA
RT   coding for a protein containing an additional 42 amino acids at the
RT   N-terminal end.";
RL   Eur. J. Biochem. 217:83-87(1993).
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY USE OF ALTERNATIVE INITIATION.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59037; CAA41761.1; -.
DR   EMBL; X74593; CAA52670.1; -.
DR   PIR; S16132; S16132.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Alternative initiation.
FT   CHAIN         1    399       SORBITOL DEHYDROGENASE, LONG ISOFORM.
FT   CHAIN        43    399       SORBITOL DEHYDROGENASE, SHORT ISOFORM.
FT   INIT_MET     43     43       FOR SHORT ISOFORM.
FT   METAL        87     87       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       112    112       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       198    198       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CONFLICT    301    301       D -> T (IN REF. 2).
SQ   SEQUENCE   399 AA;  42834 MW;  42018F127A1856D2 CRC64;

  Query Match           16.5%;  Score 331.5;  DB 1;  Length 399;
  Best Local Similarity   28.6%;  Pred. No. 3.2e-18;
  Matches  108;  Conservative   57;  Mismatches  166;  Indels   47;  Gaps   11;

Qy       4 LLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDK 63
           |: :|  |||  | | ||:  |||| :|:   ||||:|:  : :      |  |:
Db      54 LVVHGPGDIRL-ENYPIPEL-GPNDVLLKMHSVGICGSDVHYWEHGRIGDFVVKK----- 106

Qy      64 ISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCM 123
                |:  ||| :||| :||  |  :|||||||:|          | :::
Db     107 ------PMVLGHEAAGTVTKVGPMVKHLKPGDRVAIE----------PGVPREID--E 146

Qy     124 ACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHA 183
             |: |  |   |: ||      |    :  :  |   |||||:. : |||:||:||  :|
Db     147 FCKIGRYNLTPSIFFCATPPDDGNLCRFYKHSADFCYKLPDSVTFEEGALIEPLSVGIYA 206

Qy     184 VERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA----EV 239
             |     |   || | ||||: |:|  :   | ::|  : :  |   |||:||    :|
Db     207 CRRGSVSLGNKVLVCGAGPIGIVTLLVAKAMGASQVVVIDLSASRLAKAKEVGADFTIQV 266

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
            :  | |  |   :  : |      :|:|   :    | ||   |  | | : |   |
Db     267 AKETPHDIAKKVESVLGSKPE---VTIECTGAESSVQDGIYATHSGGTLVVVGM-GPEMI 322

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEK-- 357
               :|    |   |    :   | |    :::   : ::| :  |:  ||
Db     323 NLPLVHAAVREVDIKGVFRY-CNTWPMAVSMLASKTLNV---KPLVTHRFPLEKAVEAFE 378

Qy     358 ------GFKQLIEHKENN 369
                | | :|:   |:
Db     379 TAKKGLGLKVMIKCDPND 396


RESULT 14
DHSO_MOUSE
ID   DHSO_MOUSE     STANDARD;      PRT;   375 AA.
AC   Q64442;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase)
DE   (Fragment).
GN   SORD OR SDH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=95324564; PubMed=7601136;
RA   Lee F.K., Lee A.Y.W., Lin C.X.F., Chung S.S.-M., Chung S.K.;
```

```
RT   "Cloning, sequencing, and determination of the sites of expression of
RT   mouse sorbitol dehydrogenase cDNA.";
RL   Eur. J. Biochem. 230:1059-1065(1995).
CC   -!- FUNCTION: OXIDIZES SORBITOL AND OTHER POLYOLS SUCH AS XYLITOL, D-
CC       MANNITOL, L-IDITOL.
CC   -!- CATALYTIC ACTIVITY: L-iditol + NAD(+) = L-sorbose + NADH.
CC   -!- COFACTOR: ZINC.
CC   -!- PATHWAY: POLYOL PATHWAY.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative initiation (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: TESTIS HAS THE HIGHEST LEVEL OF EXPRESSION,
CC       FOLLOWED BY KIDNEY, LIVER, AND LUNG. LOW LEVELS OF EXPRESSION ARE
CC       ALSO OBSERVED IN LENS, BRAIN, AND SKELETAL MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27014; AAA79043.1; -.
DR   HSSP; P07846; 1SDG.
DR   MGD; MGI:98266; Sdh1.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Alternative initiation.
FT   NON_TER       1      1
FT   CHAIN        <1    375       SORBITOL DEHYDROGENASE, LONG ISOFORM.
FT   CHAIN        19    375       SORBITOL DEHYDROGENASE, SHORT ISOFORM.
FT   INIT_MET     19     19       FOR SHORT ISOFORM.
FT   METAL        63     63       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        88     88       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       174    174       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   375 AA;  40091 MW;  EB5912012FAA7B8C CRC64;

  Query Match             16.3%;  Score 329;  DB 1;  Length 375;
  Best Local Similarity   31.4%;  Pred. No. 4.6e-18;
  Matches   93;  Conservative   46;  Mismatches  119;  Indels   38;  Gaps    8;

Qy       4 LLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDK 63
           |: :|  |||  |  | ||:  |||| :|:   ||||:||  : :     |  |:
Db      30 LVVHGPGDIRL-ENYPIPEL-GPNDVLLKMHSVGICGSDLHYWEHGRIGDFVVKK----- 82

Qy      64 ISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCM 123
                 |:  ||| :||| :||  |  :|||||||:|             | :::
Db      83 ------PMVLGHEAAGTVTKVGELVKHLKPGDRVAIE-----------PGVPREVD--E 122

Qy     124 ACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHA 183
            |: |  |   :: ||      |    :  :  |   |||||:  : |||:||:||  :|
Db     123 YCKIGRYNLTPTIFFCATPPDDGNLCRFYKHNADFCYKLPDSVTFEEGALIEPLSVGIYA 182

Qy     184 VERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA----EV 239
             |     |   || | ||:|: |:|  :   | ::| :: :  |   |||:||    :|
Db     183 CRRGSVSLGNKVLVCGAGPVGMVTLLVAKAMGAAQVVVTDLSASRLTKAKEVGADFTIQV 242

Qy     240 FDPSTCDDANAV--LKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
             :  : |: |  |    ||        :|:|   :  : | ||   |  | | :
Db     243 GKETPQEIASKVESLLGSKPE-----VTIECTGAESSVQSGIYATHSGGTLVIVGM 293


RESULT 15
RSPB_ECOLI
ID   RSPB_ECOLI     STANDARD;      PRT;   339 AA.
AC   P38105;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Starvation sensing protein rspB (EC 1.1.1.-).
GN   RSPB OR B1580.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RX   MEDLINE=94310441; PubMed=7545940;
RA   Huisman G.W., Kolter R.;
RT   "Sensing starvation: a homoserine lactone-dependent signaling
RT   pathway in Escherichia coli.";
RL   Science 265:537-539(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- FUNCTION: NOT KNOWN; PROBABLE CATABOLIC ENZYME.
CC   -!- COFACTOR: ZINC (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31628; AAA21686.1; -.
DR   EMBL; AE000254; AAC74652.1; -.
DR   EMBL; D90799; BAA15284.1; -.
DR   EMBL; D90800; BAA15306.1; -.
DR   EcoGene; EG12452; rspB.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
FT   METAL        37     37       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        59     59       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        89     89       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        92     92       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL        95     95       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       103    103       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       144    144       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   339 AA;  36564 MW;  E0A1062DFB7C114E CRC64;

  Query Match             15.9%;  Score 319.5;  DB 1;  Length 339;
```

```
Best Local Similarity   27.2%;  Pred. No. 2.2e-17;
Matches  107;  Conservative   57;  Mismatches  159;  Indels   71;  Gaps   16;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPN--DVKIKVSYCGICGTDLKEFTYSGGPVFFPKQ 58
           || :|    | :     : | ||  |:  :|::||   ||||:|     | |   |
Db       1 MKSILIEKPNQL----AIVEREIPTPSAGEVRVKVKLAGICGSD--SHIYRGHNPF---- 50

Qy      59 GTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQD 118
             : |   |||| | :  || || | : |:||||:    |
Db      51 --------AKYPRVIGHEFFGVIDAVGEGVESARVGERVAVDPVVSC------------- 89

Qy     119 LGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPIS 178
            | |  |  | || | :|:  |:  | |||:|| |    :  |:|::: |    ::|| :
Db      90 -GHCYPCSIGKPNVCTTLAVLGV-HADGGFSEYAVVPAKNAWKIPEAVADQYAVMIEPFT 147

Qy     179 VAWHAVERARFQPGQTALVLGGGPIGLATILALQG-HHAGKIVCSEPALIRRQFAKELGA 237
           :| :     :     | || | |||||  :  |:| ::   :: ::    | : ||| ||
Db     148 IAANVTGHGQPTENDTVLVYGAGPIGLTIVQVLKGVYNVKNVIVADRIDERLEKAKESGA 207

Qy     238 EVFDPSTCDDANAVLKAMVPENEGF-------HAAFDCSGVPQTFTTSIVATGPSGIAVN 290
           :           |:  :  |  | |          | :  |     ::    |   |
Db     208 DW----------AINNSQTPLGEIFTEKGIKPTLIIDAACHPSILKEAVTLASP---AAR 254

Qy     291 VAVWG--DHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGK 348
           : : |    |  :   :| :|     |     | |  |:  |  |||   |  |:||
Db     255 IVLMGFSSEPSEVIQQGITGKELSIFSSRLNANK-FPIVIDWLSKGLI---KPEKLITHT 310

Qy     349 V---HLKDGVEKGFKQLIEHKENN-VKILVTPNE 378
             |: | :      | |  : :  |:|:| :|
Db     311 FDFQHVADAI-----SLFEQDQKHCCKVLLTFSE 339


Search completed: May  2, 2003, 12:29:29
Job time : 287 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 2, 2003, 10:19:42 ; Search time 1285 Seconds
(without alignments)
60.932 Million cell updates/sec

Title: US-10-020-674-2
Perfect score: 2013
Sequence: 1 MKGLLYYGTNDIRYSETVPE.....QLIEHKENNVKILVTPNEVS 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	588.5	29.2	349	2	Q93R65	Q93r65 bacillus ce
2	571	28.4	348	16	Q9K5Y6	Q9k5y6 bacillus ha
3	567.5	28.2	363	16	Q9HWM8	Q9hwm8 pseudomonas
4	566.5	28.1	362	2	Q59696	Q59696 pseudomonas
5	551.5	27.4	346	16	O34788	O34788 bacillus su
6	538.5	26.8	370	16	Q9CH42	Q9ch42 lactococcus
7	537	26.7	354	16	Q9K0J3	Q9k0j3 neisseria m
8	534	26.5	354	16	Q9JVJ8	Q9jvj8 neisseria m
9	455.5	22.6	352	16	Q9HWT1	Q9hwt1 pseudomonas
10	422.5	21.0	397	16	Q8U6R7	Q8u6r7 agrobacteri
11	414.5	20.6	352	5	O96496	O96496 bemisia arg
12	386.5	19.2	347	5	Q21702	Q21702 caenorhabdi
13	386.5	19.2	353	3	O74230	O74230 candida sp.
14	370	18.4	395	16	Q9WYP3	Q9wyp3 thermotoga
15	369	18.3	339	16	O35045	O35045 bacillus su
16	368.5	18.3	347	5	Q21703	Q21703 caenorhabdi
17	365.5	18.2	357	3	Q07786	Q07786 saccharomyc
18	365	18.1	371	10	Q9MAW7	Q9maw7 eriobotrya
19	364.5	18.1	348	16	Q92EF4	Q92ef4 listeria in
20	364	18.1	371	10	Q8W2D0	Q8w2d0 malus domes
21	361	17.9	347	16	Q92PZ3	Q92pz3 rhizobium m
22	361	17.9	371	10	Q9ZR22	Q9zr22 malus domes
23	358.5	17.8	348	16	Q8Y9M0	Q8y9m0 listeria mo
24	355.5	17.7	360	5	O96299	O96299 drosophila
25	352.5	17.5	354	16	Q9KGB7	Q9kgb7 bacillus ha
26	352.5	17.5	360	5	O97479	O97479 drosophila
27	352	17.5	367	10	Q9MBD7	Q9mbd7 prunus pers
28	350	17.4	368	10	Q93X81	Q93x81 prunus cera
29	349.5	17.4	360	5	Q960H1	Q960h1 drosophila
30	348	17.3	343	16	Q927H6	Q927h6 listeria in
31	348	17.3	343	16	Q8Y414	Q8y414 listeria mo
32	346	17.2	338	16	Q8ZM59	Q8zm59 salmonella
33	346	17.2	371	16	Q8UFI9	Q8ufi9 agrobacteri
34	345	17.1	358	16	Q8XDU8	Q8xdu8 escherichia
35	343	17.0	343	16	Q930C9	Q930c9 rhizobium m
36	342	17.0	289	16	Q98CF7	Q98cf7 rhizobium l
37	340	16.9	344	16	Q8XRB0	Q8xrb0 ralstonia s
38	339	16.8	343	16	Q8U8K4	Q8u8k4 agrobacteri
39	338	16.8	371	16	Q8ZBQ5	Q8zbq5 yersinia pe
40	337.5	16.8	389	16	Q8YYF3	Q8yyf3 anabaena sp
41	337	16.7	368	10	Q8W2C9	Q8w2c9 malus domes
42	336	16.7	368	10	Q8W2C8	Q8w2c8 malus domes
43	333.5	16.6	356	3	Q07993	Q07993 saccharomyc
44	333.5	16.6	357	6	O18769	O18769 callithrix
45	331.5	16.5	348	17	Q9UYX0	Q9uyx0 pyrococcus

ALIGNMENTS

RESULT 1
Q93R65
ID Q93R65 PRELIMINARY; PRT; 349 AA.
AC Q93R65;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Acetylacetoin reductase.
GN AACR.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RA Hosaka T., Ui S., Ohkuma M., Kudo T.;
RT "Characterization of the NADH-Linked Acetylacetoin Reductase / 2,3-
RT Butanediol Dehydrogenase Gene from Bacillus cereus YUF-4.";
RL J. Biosci. Bioeng. 0:0-0(2001).
DR EMBL; AB063194; BAB60856.1; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; UNKNOWN_1.
SQ SEQUENCE 349 AA; 37720 MW; 87FAA979176113E1 CRC64;

Query Match 29.2%; Score 588.5; DB 2; Length 349;
Best Local Similarity 39.4%; Pred. No. 1.6e-37;
Matches 147; Conservative 58; Mismatches 139; Indels 29; Gaps 12;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || ||:: . |:|  | |||| :| |  ||||| :|||||||| |  |  ||:| | :
Db       1 MKALLWHNQRDVRVEE-VPEPTVK-PGAVKIKVKWCGICGTDLHE--YLAGPIFIPTE-- 54

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
           :  ::  : |:  |||||| |||:| |||| | |||| ||    |                |
Db      55 EHPLTHVKAPVILGHEFSGEVVEIGEGVTSHKVGDRVVVEPIYSC--------------G 100

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
```

```
Db    101 KCEACKHGHYNVCEQLVFHGLGGEGGGFSEYTVVPEDMVHHIPDEMTYEQGALVEPAAVA 160

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE-V 239

Db    161 VHAVRQSKLKEGEAVAVFGCGPIGLLVIQAAKAAGATPVIAVELSKERQELAKLAGADYV 220

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299

Db    221 LNPAT-QDVLAEIRNLT-NGLGVNVSFEVTGVEVVLRQAIESTSFEGQTVIVSVW-EKDA 277

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359

Db    278 TITPNNLVLKEKEVVGILGYR-HIFPAVIKLISSGQI---QAEKLITKKITVDQVVEEGF 333

Qy    360 KQLIEHKENNVKI 372

Db    334 EALVKDK-TQVKI 345


RESULT 2
Q9K5Y6
ID   Q9K5Y6      PRELIMINARY;      PRT;   348 AA.
AC   Q9K5Y6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   L-iditol 2-dehydrogenase.
GN   BH3949.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AP001520; BAB07668.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   348 AA;  37516 MW;  28146770BE77FEED CRC64;

  Query Match             28.4%;  Score 571;  DB 16;  Length 348;
  Best Local Similarity   37.7%;  Pred. No. 3.6e-36;
  Matches  141;  Conservative   65;  Mismatches  134;  Indels   34;  Gaps   13;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Db      1 MQALRWHGVKDVRVDDIAEPATLK--GKVKIKVKWCGICGSDLHE--YAAGPIFIP-QNS 55

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120

Db     56 PHSLTEEKAPIVMGHEFSGQVVEVGDGVTKCEEGDRVVVEPIFAC----------G 101

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMV-KLPDSIPDDIGALVEPISV 179

Db    102 TCTACKQGKYNLCEQLGFLGLAGGGGGFSEYVTV-DEHMVHKIPDTVSFEQGALVEPAAV 160

Qy    180 AWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEV 239

Db    161 ALYAVRQSQFNVGDQAVVFGTGPIGLLVIEALKASGASKIYAVELSTERRQRAEQLGAIA 220

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299

Db    221 LNPA---EVNVVEEIQRLTDGGADVSFEVTGVPIVLTQAIESTKLNGQTMIVSIF-EKEA 276

Qy    300 GFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358

Db    277 SFQPNLVVMRERNIAGIIGY--RDIFPAVISLMNQGYFS---AETLVTKRIGLHDIVEQG 331

Qy    359 FKQLIEHKEN-NVK 371

Db    332 FEALM--KTNCNIK 343


RESULT 3
Q9HWM8
ID   Q9HWM8      PRELIMINARY;      PRT;   363 AA.
AC   Q9HWM8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   2,3-butanediol dehydrogenase.
GN   PA4153.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AE004831; AAG07540.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   363 AA;  38193 MW;  3B4DD83ACA1027B8 CRC64;

  Query Match             28.2%;  Score 567.5;  DB 16;  Length 363;
  Best Local Similarity   36.1%;  Pred. No. 7.2e-36;
  Matches  136;  Conservative   73;  Mismatches  141;  Indels   27;  Gaps   11;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Db     11 MRAAVWHGRHDIRV-EDVPLPAEPPPGWVQIRVHWCGICGSDLHE--YLAGPVFIPVE-A 66

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120

Db     67 PHPLTGLKDQCILGHEFSGEIVRLGNGVTGFAVGQAVAADACQHC----------G 112

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180

Db    113 TCYYCRHGLYNICENLAFTGLMN-NGAFAEYVNVPANLLYALPAGFPSEAGALIEPLAVG 171

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239

Db    172 MHAVKKAGSLLGQNVVVVGAGTIGLSTIMCARAAGAAQVIALEMSSARKAKALEVGASQV 231
```

```
Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
          |||  |  ||   ::|:     |   :|:|  |   |   :|  |   :|   | |  :: : |
Db    232 LDPSRC-DALGEIRALT-GGLGADVSFECIGNKHTAKLAIDAIRKAGKCVLVGIF-EEPS 288

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
            |    |   ||   |:: |   :| :|:  : || : :     ::||:: |:: ||:||
Db    289 EFNFFELVSTEKQLLGALAYN-GEFADVIAFIADGRLDI---APLVTGRIGLEEIVERGF 344

Qy    360 KQLIEHKENNVKILVTP 376
          ::|: :||:||||:|:|
Db    345 EELVNNKEHNVKIIVSP 361


RESULT 4
Q59696
ID   Q59696      PRELIMINARY;      PRT;   362 AA.
AC   Q59696;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   2,3-butanediol dehydrogenase (EC 1.1.1.4).
GN   ADH.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PPG2;
RX   MEDLINE=95113288; PubMed=7813883;
RA   Huang M., Oppermann F.B., Steinbuchel A.;
RT   "Molecular characterization of the Pseudomonas putida 2,3-butanediol
RT   catabolic pathway.";
RL   FEMS Microbiol. Lett. 124:141-150(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PPG2;
RA   Oppermann F.B.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; L35343; AAB58982.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc.
SQ   SEQUENCE   362 AA;  38386 MW;  503CD0A58C8BB816 CRC64;

  Query Match             28.1%;  Score 566.5;  DB 2;  Length 362;
  Best Local Similarity   36.6%;  Pred. No. 8.5e-36;
  Matches  138;  Conservative   65;  Mismatches  147;  Indels   27;  Gaps   11;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
            |:  :::| ||||  | || |    |  |:||| :|||||:|| |  |  ||||  | :
Db      9 MRAAVWHGRNDIRV-EQVPLPADPAPGWVQIKVDWCGICGSDLHE--YVAGPVFIPVE-A 64

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             ::| :     |||| | : ::| ||     || || :|  ||               |
Db     65 PHPLTGIQGQCILGHEFCGQIAKLGEGVEGFAVGDPVAADACQHC-------------G 110

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
          |  |  |  | |  |:| ||   :| ||| |    : : :||   | : |||:||::|
Db    111 TCYYCTHGLYNICERLAFTGLMN-NGAFAELVNVPANLLYRLPQGFPPEAGALIEPLAVG 169

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE-V 239
          |||::|    ||| :|:| | ||| ||:  :   | :::  | :  |:  |||:||  |
Db    170 MHAVKKAGSLLGQTVVVVGAGTIGLCTIMCAKAAGAAQVIALEMSSARKAKAKEVGATVV 229

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
          |||  |  ||  |  ::|:     |   :|:|  |   |   :|     :|  | |  :: : |
Db    230 LDPSQC-DALAQIRALT-FGLGADVSFECIGNKHTAKLAIDTIRKAGKCVLVGIF-EEPS 286

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
            |    |   ||   |:: |   :| :|:  : || : :    | ::||:: |:  || ||
Db    287 EFNFFELVSTEKQVLGALAYN-GEFADVIAFIADGRLDI---RPLVTGRIGLEQIVELGF 342

Qy    360 KQLIEHKENNVKILVTP 376
          ::|: :|| ||||:|:|
Db    343 EELVNNKEENVKIIVSP 359


RESULT 5
O34788
ID   O34788      PRELIMINARY;      PRT;   346 AA.
AC   O34788;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   DEHYDROGENASE.
GN   YDJL.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MARBURG 168;
RA   Kasahara Y., Nakai S., Ogasawara N., Yata K., Sadaie Y.;
RL   DNA Res. 0:0-0(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
```

```
DR   EMBL; AB007638; BAA22767.1; -.
DR   EMBL; Z99107; CAB12443.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   346 AA;  37341 MW;  5E327192D678F8A2 CRC64;

  Query Match             27.4%;  Score 551.5;  DB 16;  Length 346;
  Best Local Similarity   38.4%;  Pred. No. 1.2e-34;
  Matches  147;  Conservative   55;  Mismatches  138;  Indels   43;  Gaps   13;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||    ::   |||  | : ||: : |  ||||| :|||||:|| |  | |||:| |
Db       1 MKAARWHNQKDIRI-EHIEEPKTE-PGKVKIKVKWCGICGSDLHE--YLGGPIFIP---- 52

Qy      61 KDK---ISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEA--TSHCSDRSRYKDTV 115
           ||   ::    |:  |||||| ||||| || : | |||| ||    :|
Db      53 VDKPHPLTNETAPVTMGHEFSGEVVEVGEGVENYKVGDRVVVEPIFATH---------- 101

Qy     116 AQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVE 175
                   |: |    : | || |  |||:|||   |: : |||| :  : |||||
Db     102 ---------GHQGAYNLDEQMGFLGLAGGGGGFSEYVSVDEELLFKLPDELSYEQGALVE 152

Qy     176 PISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKEL 235
           | :|| :||  :: : |  | | | |||||  | ||:   |  |   | :  |:| |:||
Db     153 PSAVALYAVRSSKLKAGDKAAVFGCGPIGLLVIEALKAAGATDIYAVELSPERQQKAEEL 212

Qy     236 GAEVFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWG 295
           || : |||  ||  | :         |   |||: :|||     :| :|  :|  | |::|
Db     213 GAIIVDPSKTDDVVAEIAERT--GGGVDVAFEVTGVPVVLRQAIQSTTIAGETVIVSIW- 269

Qy     296 DHPIGFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDG 354
           :      |  :  :|:   | : |  :| |  |:  :::|  | |   |::| |: | |
Db     270 EKGAEIHPNDIVIKERTVKGIIGY--RDIFPAVLSLMKEGYFSAD---KLVTKKIVLDDL 324

Qy     355 VEKGFKQLIEHKENNVKILVTPN 377
           :|:||  ||: | : ||||| ||
Db     325 IEEGFGALIKEK-SQVKILVRPN 346


RESULT 6
Q9CH42
ID   Q9CH42      PRELIMINARY;      PRT;   370 AA.
AC   Q9CH42;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   2,3-butanediol dehydrogenase (EC 1.1.1.4).
GN   BUTB OR LL0897.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AE006323; AAK04995.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   370 AA;  39445 MW;  D53348B51D8030DB CRC64;

  Query Match             26.8%;  Score 538.5;  DB 16;  Length 370;
  Best Local Similarity   37.3%;  Pred. No. 1.3e-33;
  Matches  144;  Conservative   53;  Mismatches  144;  Indels   45;  Gaps   14;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           |:   :|   |||  | : || :|    | | |::|||||||| ||    ||:| |
Db      15 MRAARFYDRGDIRIDE-INEPIVK-AGQVGIDVAWCGICGTDLHEFL--DGPIFCPSAEH 70

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--ATSHCSDRSRYKDTVAQD 118
            : |:|   |:  ||| || |  :| ||: :| || | ||       :| |
Db      71 PNPITGEVPPVTLGHEMSGVVNFIGEGVSGLKVGDHVVVEPYIVPEGTDTS--------- 121

Qy     119 LGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPIS 178
                ::|  |     :| |||  || || :   |  : |:||::| |  ||:||:|
Db     122 -------ETGHYNLSEGSNFIGLGGNGGGLAEKISVDERWVHKIPDNLPLDEAALIEPLS 174

Qy     179 VAWHAVERARFQPGQTALVLGGGPIGL--ATILALQGHHAGKIVCSEPALIRRQFAKELG 236
           | :||||||      | ||:| |||||  | :   |||    :: |||: :||: |:|
Db     175 VGYHAVERANLSEKSTVLVVGAGPIGLLTAAVAKAQGH---TVIISEPSGLRRKKAQE-- 229

Qy     237 AEV----FDPSTCDDANAVLKAMVPE--NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVN 290
           |:|    |:|   |     ::| | |   :|  |||:|: |   |   : |    | |
Db     230 AQVADYFFNPIEDD-----IQAKVHEINEKGVDAAFECTSVQPGFDACLDAIRMGGTVVI 284

Qy     291 VAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVH 350
           ||:||  |       |  :|    |:: |      : :  :  | | ||   : || |:
Db     285 VAIWG-KPASVDMAKLVIKEANLLGTIAYN-NTHPKTIDLVSTGKIKLD---QFITAKIG 339

Qy     351 LKDGVEKGFKQLIEHKENNVKILVTP 376
           .  | | ::|||  || | |  |||||:|
Db     340 LDDLIDKGFDTLIHHNETAVKILVSP 365


RESULT 7
Q9K0J3
ID   Q9K0J3      PRELIMINARY;      PRT;   354 AA.
AC   Q9K0J3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Alcohol dehydrogenase, zinc-containing.
GN   NMB0604.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AE002416; AAF41031.1; -.
DR   HSSP; P07846; 1SDG.
DR   TIGR; NMB0604; -.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
```

```
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   354 AA;  37920 MW;  2EF4DBACF877C35E CRC64;

  Query Match             26.7%;  Score 537;  DB 16;  Length 354;
  Best Local Similarity   36.0%;  Pred. No. 1.6e-33;
  Matches  140;  Conservative   54;  Mismatches  147;  Indels   48;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||   :|   |||  | :||| :  |  | | |::|||||||| ||    ||:| |  |
Db       1 MKAARFYDKGDIRI-EDIPEPTVA-PGTVGINVAWCGICGTDLHEF--MEGPIFIPPCGH 56

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--------ATSHCSDRSRYK 112
              |||   |:  |||||| |  || ||  :| |  | ||         |   |:    |
Db      57 PHPISGESAPVTMGHEFSGVVYAVGEGVDDIKVGQHVVVEPYIIRDDVPTGEGSNYHLSK 116

Qy     113 DTVAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGA 172
           |                    ::| ||||  || :| :      :  : | || |  |
Db     117 D---------------------MNFIGLGGCGGGLSEKIAVKRRWVHPISDKIPLDQAA 154

Qy     173 LVEPISVAWHAVERARFQPGQTALVLGGGPIG--LATILALQGHHAGKIVCSEPALIRRQ 230
           |:||:||  ||  |:  : |  ||| | ||||  || :|  :|    |:: :| :  |: .
Db     155 LIEPLSVGHHAYVRSGAKEGDVALVGGAGPIGLLLAAVLKAKGI---KVIITELSKARKD 211

Qy     231 FAKELGAE--VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIA 288
            |:| |    :  |||  |    | |  :   ||   ||:|: | :   | : |  |:
Db     212 KARESGVADYILDPSEVDVVAEVKK--LTNGEGVDVAFECTSVNKVLDTLVEACKPAANL 269

Qy     289 VNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGK 348
           | |::|  ||      |:  :|   |:: |   |  | :| :|:| |:|:     || :
Db     270 VIVSIW-SHPATINVHSVVMKELDVRGTIAY-CNDHAETIKLVEEGKINLE---PFITQR 324

Qy     349 VHLKDGVEKGFKQLIEHKENNVKILVTPN 377
           : | : | |||::|| : |: |||:|:||
Db     325 IKLDELVSKGFERLIHNNESAVKIIVSPN 353


RESULT 8
Q9JVJ8
ID   Q9JVJ8      PRELIMINARY;      PRT;   354 AA.
AC   Q9JVJ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative zinc-binding alcohol dehydrogenase (EC 1.1.1.4).
GN   NMA0808.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AL162754; CAB84091.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   354 AA;  37888 MW;  CEE7B4CFE0C38990 CRC64;

  Query Match             26.5%;  Score 534;  DB 16;  Length 354;
  Best Local Similarity   35.7%;  Pred. No. 2.7e-33;
  Matches  139;  Conservative   55;  Mismatches  147;  Indels   48;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||   :|   |||  | :||| :  |  | | |::|||||||| ||    ||:| |  |
Db       1 MKAARFYDKGDIRI-EDIPEPTVA-PGTVGINVAWCGICGTDLHEF--MEGPIFIPPCGH 56

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--------ATSHCSDRSRYK 112
              |||   |:  |||||| |  || ||  :| |  | ||         |   |:    |
Db      57 PHPISGESAPVTMGHEFSGVVYAVGEGVDDIKVGQHVVVEPYIIHDDVPTGEGSNYHLSK 116

Qy     113 DTVAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGA 172
           |                    ::| ||||  || :| :      :  : | || |  |
Db     117 D---------------------MNFIGLGGCGGGLSEKIAVKRRWVHPISDKIPLDQAA 154

Qy     173 LVEPISVAWHAVERARFQPGQTALVLGGGPIG--LATILALQGHHAGKIVCSEPALIRRQ 230
           |:||:||  ||  |:  : |  ||| | ||||  || :|  :|    |:: :| :  |: .
Db     155 LIEPLSVGHHAYVRSGAKAGDVALVGGAGPIGLLLAAVLKAKGI---KVIITELSKARKD 211

Qy     231 FAKELGAE--VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIA 288
            |:| |    :  |||  |    | |  :   ||   ||:|: | :   | : |  |:
Db     212 KARESGVADYILDPSEVDVVEEVKK--LTNGEGVDVAFECTSVNKVLDTLVEACKPAANL 269

Qy     289 VNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGK 348
           | |::|  ||      |:  :|   |:: |   |  | :| :|:| |:|:     || :
Db     270 VIVSIW-SHPATVNVHSVVMKELDVRGTIAY-CNDHAETIKLVEEGKINLE---PFITQR 324

Qy     349 VHLKDGVEKGFKQLIEHKENNVKILVTPN 377
           : | : | :|||::|| : |: |||:|:||
Db     325 IKLDELVSEGFERLIHNNESAVKIIVSPN 353


RESULT 9
Q9HWT1
ID   Q9HWT1      PRELIMINARY;      PRT;   352 AA.
AC   Q9HWT1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Probable alcohol dehydrogenase (Zn-dependent).
GN   PA4097.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AE004826; AAG07484.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR002040; Tachykinin.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
DR   PROSITE; PS00267; TACHYKININ; UNKNOWN_1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   352 AA;  37603 MW;  8AD0F4A3AF52BBF1 CRC64;
```

```
  Query Match             22.6%;  Score 455.5;  DB 16;  Length 352;
  Best Local Similarity   33.5%;  Pred. No. 3.2e-27;
  Matches  127;  Conservative   55;  Mismatches  162;  Indels   35;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || | ::   |:| ||   | :   | :|:::|:|||||||:|| |  |  ||   | |'
Db       1 MKALRWHAARDLRLSEL--ERQAPRPGEVELEVAYCGICGSDLHE--YQSGPHSIP-QAE 55

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             :||   ||  |||| | |  :| ||   : |||||||    |               |
Db      56 AHPLSGCRAPLTLGHEFCGVVAALGPGVEGPRIGDRVAVEPEYRC-------------G 101

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
            |  |: |  | | |: | || |  |||||         : :|||::     |::|| :||
Db     102 ECRYCREGRYNLCESMGFIGLMG-DGGFAERARVPAYMLHRLPDAVGFRQAAVLEPAAVA 160

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
            ||: |:   |||   | | ||||| :: :     |   : :  |   | | ||
Db     161 LHALRRSSLAPGQRCAVFGLGPIGLLLVMLARLRGIEDIAAVDVSPERLALAGEFGASRA 220

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
            |    | |     | : |      ||: :|   :   :: :    |  | |:: |:  :
Db     221 LDARDGDTA-----ARLREGGALDCAFEAAGSQASLDAALASLRKGGELVLVSLMGE--V 273

Qy     300 GFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358
             |  :|    ||: |  :| : |::  | || : | :|   :|  | |:  || |
Db     274 RLDAFDLVNRELRLLGSVGY--RDAYPELIALLADGRLDLARA---VTRSVPLEQAVEHG 328

Qy     359 FKQLIEHKENNVKILVTPN 377
           |: |:  | : :|:|| ||
Db     329 FEALLRDK-SQLKVLVNPN 346


RESULT 10
Q8U6R7
ID   Q8U6R7      PRELIMINARY;      PRT;   397 AA.
AC   Q8U6R7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zinc-binding dehydrogenase.
GN   ATU4740 OR AGR_L_281.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009402; AAL45534.1; ALT_INIT.
DR   EMBL; AE008211; AAK88710.1; -.
KW   Complete proteome.
SQ   SEQUENCE   397 AA;  42877 MW;  C2FB48B0F9A974DF CRC64;

  Query Match             21.0%;  Score 422.5;  DB 16;  Length 397;
  Best Local Similarity   28.5%;  Pred. No. 1.4e-24;
  Matches  109;  Conservative   75;  Mismatches  160;  Indels   39;  Gaps   10;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
             |: | ::   |:|  | : ||:   |  | ::  : |||||| |::|  ||:| | :
Db      41 MRALRFHAAKDLRI-EDIAEPKRPGPGQVLVRNRFVGICGTDLHEYSY--GPIFIPTE-- 95

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE------ATSHCSDRSRYKDT 114
                :|   |   |||| | |  :| |||||  ||||:::      :  : :||  :
Db      96 PHPFTGAHGPQILGHEFGGVVEAIGDGVTSVNVGDRVSIQPLIMPRSGDYFADRGLF--- 152

Qy     115 VAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALV 174
                          :   |:  ||    || ||  :  | :: |:|| : |:  |||
Db     153 ---------------HLSTQLALVGLSWDGGGMAEAALVNEYNVQKIPDEMTDEEAALV 196

Qy     175 EPISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKE 234
           || :|| :| :|     | : || | ||||: |:|| :   | ::  |:    | : |:
Db     197 EPSAVAVYACDRGGVTAGNSVLVTGAGPIGMLTLLAARAAGATQLFVSDLNDARLELARN 256

Qy     235 LGAEV--FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVA 292
           :  :|   :|   |:   |:::       |   | :| |          : |    |: |
Db     257 VIRDVITINPKR-DNVGDVVRSATEGKVGCDVAIECVGNEHALKACVDAVRKQGVVVQTG 315

Qy     293 VW-GDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHL 351
           :   ::|| :    :|:::    ||  |    :  |:: :  ||:    | |::| :: |
Db     316 LHPHENPIDW--FQVTFRDLEIKGSWAYPTHYWPRVIRLIASGLL---PATKIVTKRITL 370

Qy     352 KDGVEKGFKQLIEHKENNVKILV 374
              | :||  |::   ::|||:
Db     371 DTAVTEGFDALLDPAGTHLKILI 393


RESULT 11
O96496
ID   O96496      PRELIMINARY;      PRT;   352 AA.
AC   O96496;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   NADP(H)-dependent ketose reductase.
GN   KR.
OS   Bemisia argentifolii (silverleaf whitefly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
OC   Aleyrodiformes; Aleyrodoidea; Aleyrodidae; Aleyrodinae; Bemisia.
OX   NCBI_TaxID=77855;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99212467; PubMed=10196734;
RA   Wolfe G.R., Smith C.A., Hendrix D.L., Salvucci M.E.;
RT   "Molecular basis for thermoprotection in Bemisia: structural
RT   differences between whitefly ketose reductase and other medium-chain
RT   dehydrogenases/reductases.";
RL   Insect Biochem. Mol. Biol. 29:113-120(1999).
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
DR   EMBL; AF067126; AAD02817.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PRINTS; PR00368; FADPNR.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   FAD; Flavoprotein; Oxidoreductase; Zinc.
```

```
SQ   SEQUENCE   352 AA;  38164 MW;  56360D3D2EDA6D26 CRC64;

  Query Match             20.6%;  Score 414.5;  DB 5;  Length 352;
  Best Local Similarity   29.0%;  Pred. No. 4.8e-24;
  Matches  113;  Conservative   65;  Mismatches  140;  Indels   71;  Gaps   10;

Qy       7 YGTNDIRYSE-TVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKIS 65
           |  ||:|  :  :|||:    ::| ::::| ||||:|:  : :     |  |
Db      11 YKQNDLRLEQRPIPEPK---EDEVLLQMAYVGICGSDVHYYEHGRIADFIVKD------- 60

Qy      66 GYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMAC 125
             |:  ||| |||||:||  |  :| |||||||    |                  | |
Db      61 ----PMVIGHEASGTVVKVGKNVKHLKKGDRVAVEPGVPCR--------------RCQFC 102

Qy     126 QSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVE 185
           : |  | |  |:||      |  | | |:  |   ||||::  : |||:||:||  ||
Db     103 KEGKYNLCPDLTFCATPPDDGNLARYYVHAADFCHKLPDNVSLEEGALLEPLSVGVHACR 162

Qy     186 RARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE---VFDP 242
           ||  | | | ||:| ||||| ::|| : : |  :||:  :  | : ||  ||:   | ||
Db     163 RAGVQLGTTVLVIGAGPIGLVSVLAAKAYGA-FVVCTARSPRRLEVAKNCGADVTLVVDP 221

Qy     243 STCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFM 302
           :  :::: : :      :  :   ||||  :  |  | |   |  : | :
Db     222 AKEEESSIIERIRSAIGDLPNVTIDCSGNEKCITIGINITRTGGTLMLVGM--------- 272

Qy     303 PMSLTYQEKYATGSMCYTVK---------DFQEVVKALEDGLISLD-------KARKMIT 346
                   ||   ||          | : | :   |  |:|:           ::::|
Db     273 -----------GSQMVTVPLVNACAREIDIKSVFRYCNDYPIALEMVASGRCNVKQLVT 320

Qy     347 GKVHLKDGVEKGFKQLIEHKENNVKILVT 375
             |:  |:  |:   :  :| :|::::
Db     321 HSFKLEQTVD-AFEAARKKADNTIKVMIS 348


RESULT 12
Q21702
ID   Q21702      PRELIMINARY;      PRT;   347 AA.
AC   Q21702;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   R04B5.5 protein.
GN   R04B5.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
DR   EMBL; Z70782; CAA94841.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc.
SQ   SEQUENCE   347 AA;  37323 MW;  AE71536D4B1344B0 CRC64;

  Query Match             19.2%;  Score 386.5;  DB 5;  Length 347;
  Best Local Similarity   29.6%;  Pred. No. 6.9e-22;
  Matches  112;  Conservative   62;  Mismatches  151;  Indels   53;  Gaps   12;

Qy       7 YGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISG 66
           || :|:|  | || |:   || | :||   ||||:|:  :|:     |  |:
Db      11 YGVDDLRL-EQVPIPK-PGPNQVLVKVHTVGICGSDVHYWTHGAIGPFVVKE-------- 60

Qy      67 YELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQ 126
             |:  ||| || | |||: |  :| |||:|:|    |                  || |:
Db      61 ---PMIVGHETSGIVSEVGNEVKHLKVGDRIAMEPGLPCK-------------LCEHCK 103

Qy     127 SGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVER 186
           :|  | |  : |       |  : :||:  |   ||||::  : |||:||:|| ||  |
Db     104 TGRYNLCPEMRFFATPPVHGTLSRFVVHDADFCFKLPDNLSFEDGALIEPLSVAIHACRR 163

Qy     187 ARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDPSTCD 246
             | |   |||| ||||:  ::  :   |||:| ::    |   ||:|||:    :| :
Db     164 GNVQMGHRVLVLGAGPIGVLNLITAKAVGAGKVVITDLDDGRLALAKKLGAD----ATIN 219

Qy     247 DANAVLKAMVPE------NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDH--- 297
               | |:  |       ::      :|:|   :  |:|  |   |:  | | :  |
Db     220 VKGKSLDAVKSEIITALGDQQPDVCIECTGAQPSIETAITTTKSGGVIVLVGLGADRVEI 279

Qy     298 PIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLK-DGVE 356
           ||    :   :|    |  | |  :   :: :  | ::|         : | | :  :
Db     280 PI----IESATREVDMRGIFRY-VNCYPTAIELISSGKLNLSGLT-----RAHYKLEETQ 329

Qy     357 KGFKQLIEHKENNVKILV 374
           : ||:    | : :|: :
Db     330 EAFKR--TQKADVIKVFI 345


RESULT 13
O74230
ID   O74230      PRELIMINARY;      PRT;   353 AA.
AC   O74230;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Xylitol dehydrogenase (EC 1.1.1.9).
GN   XDH.
OS   Candida sp. HA167.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=78167;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Habenicht A., Motejadded H., Wegerer A., Kies M., Mattes R.;
RT   "Xylose utilisation: Cloning and characterisation of the xylitol
RT   dehydrogenase from Galactocandida mastotermitis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
DR   EMBL; AF072541; AAC24597.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc.
SQ   SEQUENCE   353 AA;  37382 MW;  9BA7F10CE9F6E687 CRC64;

  Query Match             19.2%;  Score 386.5;  DB 3;  Length 353;
  Best Local Similarity   30.6%;  Pred. No. 7.1e-22;
  Matches  114;  Conservative   48;  Mismatches  161;  Indels   49;  Gaps    9;

Qy      11 DIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELP 70
           |::: |  | |:: :|  |||:|   ||||:|:  ||:     |  |           |
Db      16 DVKF-EDRPIPKLSDPYSVKIQVKKTGICGSDVHYFTHGAIGDFVVK----------AP 63

Qy      71 LCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSP 130
           :  ||| || |:|||| | |:| |||||:|          ||              ||
Db      64 MVLGHESSGVVLEVGSEVKSLKVGDRVAMEPGVPSRHSDEYK-------------SGRY 109

Qy     131 NCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQ 190
```

```
Db    110 NLCPHMAFAATPPYDGTLCKYYILPEDFCVKLPEHVSLEEGALVEPLSVAVHSSKLGNIK 169

Qy    191 PGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEV-----FDPSTC 245
Db    170 PGSHVAIYGAGPVGLLVAAVASAFGAESVTIIDLVESRLNLAKELGATATVQVDFKDTPK 229

Qy    246 DDANAVLKAMVPENEGF--HAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMP 303
Db    230 ESAAKVVAA----NNGIAPDVVIDASGAEASINSAINAIRPGGTYVQVGM-GKPDVSFPI 284

Qy    304 MSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVE------- 356
Db    285 ATLIGKELTVKGSFRYGYGDYPLAVSLLASGKVNV---KKLITHEVKFEDAAEAFQLVRD 341

Qy    357 -KGFKQLIEHKE 367
Db    342 GKAIKCIINGPE 353


RESULT 14
Q9WYP3
ID   Q9WYP3      PRELIMINARY;      PRT;   395 AA.
AC   Q9WYP3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Alcohol dehydrogenase, zinc-containing.
GN   TM0412.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AE001720; AAD35497.1; -.
DR   TIGR; TM0412; -.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   395 AA;  43317 MW;  29D2A19F00946397 CRC64;

  Query Match             18.4%;  Score 370;  DB 16;  Length 395;
  Best Local Similarity   34.0%;  Pred. No. 1.6e-20;
  Matches  109;  Conservative   37;  Mismatches  111;  Indels   64;  Gaps   13;

Qy      4 LLYYGTNDIRYSET----VPEPEIKNPNDVKIKVSYCGICGTDL-KEFTYSGGPVFFPKQ 58
Db     27 LTWLGSKVWRYPEVRVEEVPEPRIEKPTEIIIKVKACGICGSDVHMAQTDEEGYILYP-- 84

Qy     59 GTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKP------GDRVAVEATSHCSDRSRYK 112
Db     85 ----GLTGF--PVTLGHEFSGVVVEAGPEAINRRTNKRFEIGEPVCAEEMLWC------- 131

Qy    113 DTVAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYV--------------VYGED 157
Db    132 -------GHCRPCAEGFPNHCENLNELGF-NVDGAFAEYVKVDAKYAWSLRELEGVYEGD 183

Qy    158 HMVKLPDSIPDDIGALVEPISVAWHA--VERARFQPGQTALVLGGGPIGLATILALQGHH 215
Db    184 RLFL--------AGSLVEPTSVAYNAVIVRGGGIRPGDNVVILGGGPIGLAAVAILKHAG 235

Qy    216 AGKIVCSEPALIRRQFAKELGAE-VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQT 274
Db    236 ASKVILSEPSEVRRNLAKELGADHVIDPTKENFVEAVLD--YTNGLGAKLFLEATGVPQL 293

Qy    275 FTTSI--VATGPSGIAVNVAV 293
Db    294 VWPQIEEVIWRARGINATVAI 314


RESULT 15
O35045
ID   O35045      PRELIMINARY;      PRT;   339 AA.
AC   O35045;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Zinc-containing alcohol dehydrogenase.
GN   YJMD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Rivolta C., Soldo B., Lazarevic V., Joris B., Mauel C., Karamata D.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AF015825; AAC46329.1; -.
DR   EMBL; Z99110; CAB13090.1; -.
DR   HSSP; P07846; 1SDG.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   Pfam; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ   SEQUENCE   339 AA;  36762 MW;  BD91891649037700 CRC64;

  Query Match             18.3%;  Score 369;  DB 16;  Length 339;
  Best Local Similarity   30.3%;  Pred. No. 1.5e-20;
  Matches  115;  Conservative   56;  Mismatches  162;  Indels   46;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || :      |:  :| | :| :   ::| :||   ||||:|:  :           ||
Db       1 MKAVQVRKAYDLVTAE-VKKPVLSKDDEVLVKVKRVGICGSDMHIY----------HGT 48

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
              :   ||   ||| :| |  ||: | |:|||| | :|  |:|               |
Db      49 NPLAT---LPRVIGHEVTGQVEAVGANVQSLKPGDHVVIEPISYC--------------G 91

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           | ||: | || || ||  |:    ||  ||:|  |  :  :   :| :   : || ::
Db      92 SCYACRKGRPNVCAKLSVFGV-HEDGGMREYIVLPERQLHAVSKDLPWEEAVMAEPYTIG 150

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE-V 239
             || | : : | | |: | ||||:  :| :       :: ::    |  |||| ||: |
Db     151 AQAVWRGQVEKGDTVLIQGAGPIGIC-VLKMAKLAGAAVMMTDLNNERLAFAKENGADAV 209

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
            :      |  ||:     ||| :   |   :|:||  || |  |:|  | |  : :
Db     210 VNVQAEHVAERVLE--WTGNEGANVVIDAVCLPETFALSIEAVSPAGHVV-VLGFDERAA 266

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKV---HLKDGVE 356
           : :| :|   |||   |   | :||: |  |       | |  | |      | |
Db     267 QISQLPITKKEVTITGSRLQT-NQFPKVVELLNGG-------RLMHNGLVTHTFSVDDVH 318

Qy     357 KGFKQLIEHKENNVKILVT 375
             |: : || :   | ::|
Db     319 HAFQFIKEHPDQVRKAVIT 337


Search completed: May  2, 2003, 12:51:08
Job time : 1287 secs
```

OM protein - protein search, using sw model

Run on: May 2, 2003, 09:32:12 ; Search time 939 Seconds
(without alignments)
53.925 Million cell updates/sec

Title: US-10-020-674-2
Perfect score: 2013
Sequence: 1 MKGLLYYGTNDIRYSETVPE.....QLIEHKENNVKILVTPNEVS 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	567.5	28.2	363	22	AAU36386	Pseudomonas aerugi
2	538.5	26.8	370	23	ABB54213	Lactococcus lactis
3	519.5	25.8	345	23	AAE21046	Human drug metabol
4	491.5	24.4	354	23	ABP38529	Staphylococcus epi
5	455.5	22.6	352	22	AAU36382	Pseudomonas aerugi
6	417.5	20.7	291	22	AAG81823	S. epidermidis ope
7	360.5	17.9	357	15	AAR56479	Sorbitol dehydroge
8	358.5	17.8	348	23	ABB47897	Listeria monocytog
9	357	17.7	350	22	AAE05988	Clostridium diffic
10	355.5	17.7	360	22	ABB60809	Drosophila melanog
11	352.5	17.5	354	22	AAE05985	Bacillus haloduran
12	352.5	17.5	360	22	ABB58718	Drosophila melanog
13	349	17.3	316	22	AAB47473	G. max sorbitol de
14	348	17.3	343	23	ABB49510	Listeria monocytog
15	346	17.2	368	23	AAM47254	Arxula adeninivora
16	342	17.0	358	22	AAU34594	E. coli cellular p
17	342	17.0	363	12	AAR14181	Xylitol dehydrogen
18	337.5	16.8	343	22	AAE05994	Bacillus haloduran
19	337.5	16.8	350	22	AAE05986	Clostridium diffic
20	336	16.7	343	23	AAU77018	Idonic acid dehydr
21	334	16.6	352	22	AAE05987	Clostridium diffic
22	331.5	16.5	348	22	AAB96816	Putative P. abyssi
23	329.5	16.4	348	21	AAY82330	Pyrococcus horikos
24	328	16.3	352	21	AAG47840	Arabidopsis thalia
25	328	16.3	365	21	AAG47839	Arabidopsis thalia
26	328	16.3	370	21	AAG47838	Arabidopsis thalia
27	327	16.2	301	22	AAB47474	T. aestivum sorbit
28	327	16.2	346	21	AAG13242	Arabidopsis thalia
29	327	16.2	359	21	AAG13241	Arabidopsis thalia
30	327	16.2	364	21	AAG13240	Arabidopsis thalia
31	326.5	16.2	356	21	AAB56748	Human prostate can
32	326	16.2	341	22	AAB47471	Z. mays sorbitol d
33	323	16.0	346	21	AAG19409	Arabidopsis thalia
34	323	16.0	359	21	AAG19408	Arabidopsis thalia
35	323	16.0	364	21	AAG19407	Arabidopsis thalia
36	319	15.8	343	22	AAU34209	Staphylococcus aur
37	319	15.8	354	22	AAU36871	Staphylococcus aur
38	315	15.6	347	22	AAU38180	Salmonella typhi c
39	314	15.6	352	22	AAE05993	Enterococcus avium
40	308	15.3	341	22	AAU38240	Salmonella typhi c
41	307.5	15.3	346	22	AAU34613	E. coli cellular p
42	307.5	15.3	346	22	AAG98373	Escherichia coli p
43	306	15.2	342	22	AAU35365	Haemophilus influe
44	303	15.1	336	22	ABB52481	Escherichia coli p
45	300	14.9	317	21	AAG47846	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAU36386
ID AAU36386 standard; Protein; 363 AA.
XX
AC AAU36386;
XX
DT 14-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa cellular proliferation protein #376.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

```
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54245.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 11979; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   363 AA;

  Query Match             28.2%;  Score 567.5;  DB 22;  Length 363;
  Best Local Similarity   36.1%;  Pred. No. 7.5e-48;
  Matches  136;  Conservative   73;  Mismatches  141;  Indels   27;  Gaps   11;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           |:  :::| :|||  | || |    |  |:|:| :|||||:|| |  |  |||| | :
Db      11 MRAAVWHGRHDIRV-EDVPLPAEPPPGWVQIRVHWCGICGSDLHE--YLAGPVFIPVE-A 66

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             ::| :     ||||| :| :|:|||    |  || :|  ||                |
Db      67 PHPLTGLKDQCILGHEFSGEIVRLGNGVTGFAVGQAVAADACQHC-------------G 112

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           |  |: |  | | :|:| ||   :| |||||    : :  ||   | : |||:||::|
Db     113 TCYYCRHGLYNICENLAFTGLMN-NGAFAEYVNVPANLLYALPAGFPSEAGALIEPLAVG 171

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
           |||:|    ||  :|:| | |||:||:  :   | :::  | :  |:  | |:|| :|
Db     172 MHAVKKAGSLLGQNVVVVGAGTIGLSTIMCARAAGAAQVIALEMSSARKAKALEVGASQV 231

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           |||  |  ||   ::|:     |   :|:| |   |   :| |   :|  | | :: : |
Db     232 LDPSRC-DALGEIRALT-GGLGADVSFECIGNKHTAKLAIDAIRKAGKCVLVGIF-EEPS 288

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
           |  .|   ||   |:: |   :| :|:  : || : :     ::||:: |:: ||:||
Db     289 EFNFFELVSTEKQLLGALAYN-GEFADVIAFIADGRLDI---APLVTGRIGLEEIVERGF 344

Qy     360 KQLIEHKENNVKILVTP 376
           ::|: :||:|||||:|:|
Db     345 EELVNNKEHNVKIIVSP 361


RESULT 2
ABB54213
ID   ABB54213 standard; Protein; 370 AA.
XX
AC   ABB54213;
XX
DT   16-MAY-2002  (first entry)
XX
DE   Lactococcus lactis protein butB.
XX
KW   Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS   Lactococcus lactis IL1403.
XX
PN   FR2807446-A1.
XX
PD   12-OCT-2001.
XX
PF   11-APR-2000; 2000FR-0004630.
XX
PR   11-APR-2000; 2000FR-0004630.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Bolotine A,  Sorokine A,  Renault P,  Ehrlich SD;
XX
DR   WPI; 2002-043418/06.
XX
PT   New nucleotide sequence useful in the identification or Lactococcus
PT   lactis and related species -
XX
PS   Claim 6; SEQ ID No 915; 2504pp; French.
XX
CC   The present invention is related to a Lactococcus lactis nucleotide
CC   sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC   nucleic acid sequence is useful in the detection and/or amplification of
CC   nucleic acid sequence, particularly to identify Lactococcus lactis or
CC   related species. The proteins of the invention are useful for the
CC   biosynthesis or biodegradation of a composition of interest. The
CC   invention helps research in lactic bacteria, particularly useful in the
CC   production of yogurt and cheese.
CC   Note: The sequence data for this patent is based on equivalent patent
CC   WO200177334 (published 18-OCT-2001) which is available in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   370 AA;

  Query Match             26.8%;  Score 538.5;  DB 23;  Length 370;
  Best Local Similarity   37.3%;  Pred. No. 6.2e-45;
  Matches  144;  Conservative   53;  Mismatches  144;  Indels   45;  Gaps   14;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           |:   :|   |||  | : || :|    | | |::|||||||| ||    ||:| |
Db      15 MRAARFYDRGDIRIDE-INEPIVK-AGQVGIDVAWCGICGTDLHEFL--DGPIFCPSAEH 70

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--ATSHCSDRSRYKDTVAQD 118
           : |:|   |:  ||| || |  :| ||: :| || | ||       :| |
Db      71 PNPITGEVPPVTLGHEMSGVVNFIGEGVSGLKVGDHVVVEPYIVPEGTDTS--------- 121

Qy     119 LGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPIS 178
                 ::|  |     :| |||| || || :   |  : |:||::| |  ||:||:|
Db     122 -------ETGHYNLSEGSNFIGLGGNGGGLAEKISVDERWVHKIPDNLPLDEAALIEPLS 174

Qy     179 VAWHAVERARFQPGQTALVLGGGPIGL--ATILALQGHHAGKIVCSEPALIRRQFAKELG 236
           | :||||||      | ||:| |||||  | :   |||    :: |||: :||: |:|
Db     175 VGYHAVERANLSEKSTVLVVGAGPIGLLTAAVAKAQGH---TVIISEPSGLRRKKAQE-- 229

Qy     237 AEV----FDPSTCDDANAVLKAMVPE--NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVN 290
           |:|    |:|   |      ::| | |   :|  |||:|: |   |   : |    |  |
Db     230 AQVADYFFNPIEDD-----IQAKVHEINEKGVDAAFECTSVQPGFDACLDAIRMGGTVVI 284

Qy     291 VAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVH 350
           ||:||  |       |  :|    |:: |      : :  :  | | ||   : || |:
Db     285 VAIWG-KPASVDMAKLVIKEANLLGTIAYN-NTHPKTIDLVSTGKIKLD---QFITAKIG 339

Qy     351 LKDGVEKGFKQLIEHKENNVKILVTP 376
           | | ::|||  || | |  |||||:|
Db     340 LDDLIDKGFDTLIHHNETAVKILVSP 365


RESULT 3
AAE21046
```

ID AAE21046 standard; Protein; 345 AA.
XX
AC AAE21046;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human drug metabolising enzyme (DME-4) protein.
XX
KW Human; drug metabolising enzyme; cell proliferative disorder; metabolic;
KW autoimmune; inflammatory; developmental; gastrointestinal; hypergonadal;
KW pancreatic; endocrine; eye; dermatitis; Addison's disease; antilipaemic;
KW acquired immunodeficiency syndrome; AIDS; glomerulonephritis; anorectic;
KW diabetes; atherosclerosis; adult respiratory distress syndrome; anaemia;
KW Grave's disease; thyroiditis; Crohn's disease; infection; anticoagulant;
KW systemic lupus erythematosus; cirrhosis; psoriasis; epilepsy; gastritis;
KW cataract; hypopituitarism; cancer; rheumatoid arthritis; conjunctivitis;
KW cystic fibrosis; peptic ulcer; Wilson's disease; hepatitis; antithyroid;
KW allergy; diarrhoea; thrombosis; obesity; immunostimulant; tranquillizer;
KW infertility; vulnerary; anticonvulsant; gynaecological; laxative; goitre;
KW nootropic; jaundice; trauma; asthma; DME-4; enzyme.
XX
OS Homo sapiens.
XX
PN WO200212467-A2.
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-US24382.
XX
PR 04-AUG-2000; 2000US-223055P.
PR 11-AUG-2000; 2000US-224728P.
PR 18-AUG-2000; 2000US-226440P.
PR 24-AUG-2000; 2000US-228067P.
PR 31-AUG-2000; 2000US-230063P.
PR 13-SEP-2000; 2000US-232244P.
PR 20-SEP-2000; 2000US-234269P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Baughn MR, Bruns CM, Das D, Delegeane AM, Ding L, Elliot VS;
PI Gandhi AR, Griffin JA, Hafalia AJA, Khan FA, Lal P, Lee S;
PI Lu DAM, Lu Y, Patterson C, Ramkumar J, Ring HZ, Sanjanwala MS;
PI Tang YT, Thangavelu K, Thornton M, Tribouley CM, Walia NK;
PI Warren BA, Yang J, Yao MG, Yue H;
XX
DR WPI; 2002-206331/26.
DR N-PSDB; AAD33483.
XX
PT New human drug metabolizing enzyme polypeptide and polynucleotide
PT useful for diagnosing, treating and preventing cell proliferative,
PT autoimmune/inflammatory, endocrine, eye, metabolic and gastrointestinal
PT disorders -
XX
PS Claim 48; Page 149-150; 179pp; English.
XX
CC The invention relates to an isolated human drug metabolising enzyme (DME)
CC polypeptide or a biologically active or immunogenic fragment of DME. DME
CC is useful for diagnosis, treatment and prevention of cell proliferative,
CC autoimmune/inflammatory, developmental, endocrine, eye, metabolic and
CC gastrointestinal disorders including live disorders. Autoimmune/
CC inflammatory disorders include acquired immunodeficiency syndrome (AIDS),
CC adult respiratory distress syndrome, Addison's disease, atherosclerosis,
CC allergies, anaemia, asthma, autoimmune haemolytic anaemia, autoimmune
CC thyroiditis, Crohn's disease, atopic dermatitis, diabetes mellitus,
CC glomerulonephritis, rheumatoid arthritis, systemic lupus erythematosus,
CC ulcerative colitis, uveitis, viral, bacterial, protozoal, parasitic,
CC fungal, helminthic infections and trauma. Cell proliferative disorders
CC include cancer, arteriosclerosis, cirrhosis and psoriasis; developmental
CC disorders include epilepsy and cataract; and endocrine disorders include
CC disorders of hypothalamus/pituitary, disorders associated with
CC hypopituitarism, including diabetes insipidus, hypogonadism, disorders
CC associated with hypothyroidism including goitre, Grave's disease,
CC pancreatic disorders such as diabetes mellitus, disorders associated with
CC adrenals, disorders associated with gonadal steroid hormones such as
CC endometriosis, infertility, hypergonadal disorders and gynaecomastia.
CC Disorders of the eye include conjunctivitis and macular degeneration and
CC metabolic disorders include diabetes, cystic fibrosis, obesity and
CC hypocalcaemia. Gastrointestinal disorders include gastritis, peptic
CC ulcer, hepatitis, constipation, diarrhoea, jaundice, Wilson's disease,
CC thrombosis and hepatic tumours. DME gene is useful in gene therapy. The
CC present sequence is human DME-4 protein.
XX
SQ Sequence 345 AA;

Query Match 25.8%; Score 519.5; DB 23; Length 345;
Best Local Similarity 35.4%; Pred. No. 4.4e-43;
Matches 134; Conservative 62; Mismatches 142; Indels 41; Gaps 12;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||  ::||  |:|  |   ||:    |:||:|||: ||||||| |  |  ||:|   :
Db       1 MKAAVWYGQKDVRVEER--EPKELQDNEVKVKVSWAGICGTDLHE--YLEGPIFISTE-K 55

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             |   | : |:  ||||:| | | || ||    |||| |  |   | | : ::    |
Db      56 PDPFLGQKAPVTLGHEFAGVVEETGSQVTKFNKGDRVVVNPT--VSKREKEENIDLYD-- 111

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
                          || || |: |||||:    |::: ||||:: |  |||||| :||
Db     112 --------------GYSFIGL-GSDGGFAEFTNAPEENVYKLPDNVSDKEGALVEPTAVA 156

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
             |::      | |  : | ||||| |::| :   | ||   : :  |   || |||
Db     157 VQAIKEGEVLFGDTVAIFGAGPIGLLTVVAAKAAGASKIFVFDLSEERLSKAKALGATHA 216

Qy     240 FDPSTCDDANAVLKAMVPENE----GFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWG 295
           :    |  : :        ||    |   :|: :||  |  :||  |   |  | |:::|
Db     217 INSGKTDPVDVI-------NEYTENGVDVSFEVAGVAPTLKSSIDVTKARGTVVIVSIFG 269

Qy     296 DHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGV 355
            ||| : || ||       | :: ||   ||: :  : :| :::   : ::| :: |:: |
Db     270 -HPIEWNPMQLTNTGVKLTSTIAYTPTTFQQTIDLINEGNLNV---KDVVTDEIELENIV 325

Qy     356 EKGFKQLIEHKENNVKILV 374
           | ||:||:  | :  |||:
Db     326 ESGFEQLVNDK-SQAKILI 343
```

RESULT 4
ABP38529
ID ABP38529 standard; Protein; 354 AA.
XX
AC ABP38529;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3374.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.

```
DR   N-PSDB; ABN91074.
XX
PT   Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT   polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS   Disclosure; SEQ ID 3374; 267pp; English.
XX
CC   ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC   frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC   given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC   antibacterial activity and can be used in gene therapy. The sequences
CC   can also be used in the diagnosis and treatment of bacterial infections,
CC   particularly S. epidermidis infections. The sequences can be used to
CC   screen for compounds able to interfere with the S. epidermidis life
CC   cycle or inhibit S. epidermidis infection.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   USPTO web site.
XX
SQ   Sequence   354 AA;

  Query Match             24.4%;  Score 491.5;  DB 23;  Length 354;
  Best Local Similarity   34.9%;  Pred. No. 2.9e-40;
  Matches  132;  Conservative   59;  Mismatches  154;  Indels   33;  Gaps   11;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||  ::||  |:|  :   ||:    |:|::|||: |||||| |  |  ||:|  .
Db       5 MKAAVWYGQKDVRVEDR--EPKAIKDNEVQVKVSWAGICGTDLHE--YLEGPIFI-STDQ 59

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
            | : |   |:  |||||| |  ||  |:  | |||| |  |   | | : ::    |
Db      60 PDPLLGQTAPVTLGHEFSGVVENVGKNVSRFKKGDRVVVNPT--VSKREKPENVDLYD-- 115

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
                       || || |: | |||:    | ::  |||::    |||||| :||
Db     116 -------------GYSFIGL-GSDGAFAEFTNAPETNVYHLPDNVSAREGALVEPTAVA 160

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
             ||:      | |  | | ||||| ||:| :   | ||   : :  |   ||| :||  |
Db     161 VQAVKEGELLFGDTVAVFGAGPIGLLTIVAAKAAGASKIFVFDLSEERLAKAKSVGATHV 220

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           ::    |    | :    : |   :|: :|| |   ||  | | | || |:::| ||:
Db     221 YNSGNVDPVQTVYE---HTDNGVDVSFEVAGVGITLQQSIEVTRPRGTAVIVSIFG-HPV 276

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
           | |:    :    | :: ||   ||: :  : :| :::   : ::| :: | : || ||
Db     277 EFNPLLQMNKGVKLTTTIAYTPTTFQQTIDLIANGSLNV---KDVVTDQIELDNIVESGF 333

Qy     360 KQLIEHKENNVKILVTPN 377
             ||:  | :  |||| |
Db     334 NQLVNDK-SQAKILVRLN 350


RESULT 5
AAU36382
ID   AAU36382 standard; Protein; 352 AA.
XX
AC   AAU36382;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Pseudomonas aeruginosa cellular proliferation protein #372.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54241.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 11975; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   352 AA;

  Query Match             22.6%;  Score 455.5;  DB 22;  Length 352;
  Best Local Similarity   33.5%;  Pred. No. 1.2e-36;
  Matches  127;  Conservative   55;  Mismatches  162;  Indels   35;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || | ::   |:| ||   | :   | :|:::|:||||||:|| |  |  ||   | |
Db       1 MKALRWHAARDLRLSEL--ERQAPRPGEVELEVAYCGICGSDLHE--YQSGPHSIP-QAE 55

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
              :||   ||  |||| | |  :| ||   : |||||||    |              |
Db      56 AHPLSGCRAPLTLGHEFCGVVAALGPGVEGPRIGDRVAVEPEYRC-------------G 101

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
             |  |: |  | | |: | || |  |||||        : :|||::     |::|| :||
Db     102 ECRYCREGRYNLCESMGFIGLMG-DGGFAERARVPAYMLHRLPDAVGFRQAAVLEPAAVA 160

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
             ||: |:   |||   | | |||||  ::  :       |   : :  |   | | ||
Db     161 LHALRRSSLAPGQRCAVFGLGPIGLLLVMLARLRGIEDIAAVDVSPERLALAGEFGASRA 220

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
             |    | |     | : |      ||: :|   :  :: :    |  | |:: |:  :
Db     221 LDARDGDTA-----ARLREGGALDCAFEAAGSQASLDAALASLRKGGELVLVSLMGE--V 273

Qy     300 GFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358
                 |  :|    ||: |  :| : |::  | || : | :|   :|  | |:  || |
Db     274 RLDAFDLVNRELRLLGSVGY--RDAYPELIALLADGRLDLARA---VTRSVPLEQAVEHG 328
```

```
Qy    359 FKQLIEHKENNVKILVTPN 377
          |: |:  | : ::|:|| ||
Db    329 FEALLRDK-SQLKVLVNPN 346
```

RESULT 6
AAG81823
ID AAG81823 standard; Protein; 291 AA.
XX
AC AAG81823;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:740.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
DR N-PSDB; AAH52673.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 18; Page 228; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 291 AA;

```
  Query Match             20.7%;  Score 417.5;  DB 22;  Length 291;
  Best Local Similarity   36.4%;  Pred. No. 5.4e-33;
  Matches  111;  Conservative   42;  Mismatches  123;  Indels   29;  Gaps    9;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
            ||  ::||  |:|  :   ||:    |:|::|||: ||||||| |  |  ||:|
Db      1 MKAAVWYGQKDVRVEDR--EPKAIKDNEVQVKVSWAGICGTDLHE--YLEGPIFI-STDQ 55

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
          | : |   |:  |||||| :  ||  |:  | |||| |  |   | | : ::    |

Db     56 PDPLLGQTAPVTLGHEFSGVIENVGKDVSRFKKGDRVVVNPT--VSKREKPENVDLYD-- 111

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
                          || || |: | |||:    | ::  |||::    |||||| :||
Db    112 --------------GYSFIGL-GSDGAFAEFTNAPETNVYHLPDNVSAREGALVEPTAVA 156

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
            ||:      | |  | | ||||| ||:| :   | ||   : :  |   || :|| |
Db    157 VQAVKEGELLFGDTVAVFGAGPIGLLTIVAAKAAGASKIFVFDLSEERLAKAKSVGATHV 216

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
            ::    |    | :    : |   :|: :||  |   ||  | | | || |:::| ||:
Db    217 YNSGNVDPVQTVYE---HTDNGVDVSFEVAGVGITLQQSIEVTRPRGTAVIVSIFG-HPV 272

Qy    300 GFMPM 304
            | |:
Db    273 EFNPL 277
```

RESULT 7
AAR56479
ID AAR56479 standard; protein; 357 AA.
XX
AC AAR56479;
XX
DT 07-MAR-1995 (first entry)
XX
DE Sorbitol dehydrogenase from Saccharomyces cerevisiae.
XX
KW Sorbitol dehydrogenase; yeast regulatory sequence; assay reagent;
KW heterologous protein expression; superoxide dismutase; antibody;
KW hepatitis delta antigen.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9415942-A.
XX
PD 21-JUL-1994.
XX
PF 22-DEC-1993; 93WO-US12560.
XX
PR 30-DEC-1992; 92US-0998226.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Sarthy AV, Schopp CW;
XX
DR WPI; 1994-249120/30.
DR N-PSDB; AAQ66579.
XX
PT Purified yeast sorbitol dehydrogenase - used for isolating gene
PT regulatory sequences for use in expressing heterologous
PT polypeptide(s) in yeast
XX
PS Claim 27; Page 31-34; 60pp; English.
XX
CC This sequence (AAR56479) is that of sorbitol dehydrogenase.
CC The presence of eleven amino acids of the sorbitol dehydrogenase
CC from its N-terminal in a fusion polypeptide results in several fold
CC higher levels of expression of the fused polypeptides as compared to
CC the unfused polypeptide. The DNA encoding the enzyme is shown in
CC AAQ66579, it is useful for its regulatory elements. These can be used
CC in yeast hosts for the production of heterologous proteins, eg.
CC superoxide dismutase or the hepatitis delta antigen.
XX
SQ Sequence 357 AA;

```
  Query Match             17.9%;  Score 360.5;  DB 15;  Length 357;
  Best Local Similarity   30.3%;  Pred. No. 3.8e-27;
  Matches  111;  Conservative   56;  Mismatches  160;  Indels   39;  Gaps   11;

Qy     16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYEL--PLCP 73
          | | | ||:|: ||: :   ||||:|:  :  |||           | |_|  |:
```

```
Db      20 EQRPIPTIKDPHYVKLAIKATGICGSDI-HYYRSGG-----------IGKYILKAPMVL 66

Qy      74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
           ||| || ||||| || || |||||:|        ||| |           : |  | |
Db      67 GHESSGQVVEVGDAVTRVKVGDRVAIEP----GVPSRYSDET----------KEGRYNLC 112

Qy     134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
             ::|       |   :| :  || :||||: :  : || |||:||  |: : |  : |
Db     113 PHMAFAATPPIDGTLVKYYLSPEDFLVKLPEGVSYEEGACVEPLSVGVHSNKLAGVRFGT 172

Qy     194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPS--TCDDANA 250
             :| | ||:|| |    :   |  ::  :    : | ||: ||   |: |  : | |
Db     173 KVVVFGAGPVGLLTGAVARAFGATDVIFVDVFDNKLQRAKDFGATNTFNSSQFSTDKAQD 232

Qy     251 VLKAMVPENEGFHA--AFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTY 308
           :   :    | ||   |:|||       ::  |   |  | | : | :   |   ::
Db     233 LADGVQKLLGGNHADVVFECSGADVCIDAAVKTTKVGGTMVQVGM-GKNYTNFPIAEVSG 291

Qy     309 QEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKEN 368
           :|    |   |:  |::: |  :  | :::   : :|| |   :|   | :   | |
Db     292 KEMKLIGCFRYSFGDYRDAVNLVATGKVNV---KPLITHKFKFEDAA-KAYDYNIAHGGE 347

Qy     369 NVKILV 374
             || ::
Db     348 VVKTII 353


RESULT 8
ABB47897
ID   ABB47897 standard; Protein; 348 AA.
XX
AC   ABB47897;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes protein #601.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C,  Frangeul L,  Couve E,  Rusniok C,  Fsihi H,  Dehoux P;
PI   Dussurget O,  Chetouani F,  Nedjari H,  Glaser P,  Kunst F,  Cossart P;
PI   Daniels J,  Goebel W,  Kreft J,  Kuhn M,  Ng E,  Vazquez-Boland JA;
PI   Dominguez-Bernal G,  Garrido-Garcia P,  Tierrez-Martinez A,  Amend A;
PI   Chakraborty T,  Domann E,  Hain T,  Berche P,  Charbit A,  Durant L;
PI   Perez-Diaz J,  Baquero F,  Garcia Del Portillo F,  Gomez-Lopez N;
PI   Maduenio E,  De Pablos B,  Wehland J,  Kaerst U,  Entian K,  Hauf J;
PI   Rose M,  Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides  -
XX
PS   Claim 6; SEQ ID No 602; 192pp; French.
XX
CC   The present invention relates to the genome sequence of Listeria
CC   monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC   it are useful for selecting probes and primers for detecting genes in L.
CC   monocytogenes and related organisms, and for studying genetic
CC   polymorphisms and other genomes. The present sequence is a protein
CC   encoded by the genome sequence of the present invention. Proteins
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   348 AA;

  Query Match             17.8%;  Score 358.5;  DB 23;  Length 348;
  Best Local Similarity   28.3%;  Pred. No. 5.7e-27;
  Matches  110;  Conservative   65;  Mismatches  159;  Indels   55;  Gaps   11;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || |  ||  |:|| |     | |: |:|| :||   ||||:|:  ::
Db       1 MKALKLYGKRDLRYEE-ADMPTIEEPDDVILKVKTVGICGSDISRYS---------- 46

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             |:  |   :  |||||| |:|||| || :: ||| |     :|           |
Db      47 --KLGPYVPGMVWGHEFSGEVIEVGSEVTDIEIGDRAAGCPALYC----------G 90

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
             |  |: |    |  |:  |     | :|||:    :::||:|: :  :  ||||| :|
Db      91 ECEYCKKGEFARCRKLTVIG-ARHPGAYAEYIKLPAENVVKIPNELDYEAAALVEPSAVV 149

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
             |     : | |   :|:| | |||  |   :  | :::  :    :   |||:||:|
Db     150 VHGFYHTKLQAGDDVVVVGSGNIGLLAIQWAKVFGARRVIAIDVDDKKLALAKEVGADVV 209

Qy     241 DPSTCDDANAVLKAMVPENEGFHA--AFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHP 298
           |  :|    |: :    :| :|    : :| | | :  : |    |  |   |:  |
Db     210 INSLKEDP---LEVVAAHTDGLNADLVVEAAGSPIT-SAQVFAYAKKGGGV---VFLGIP 262

Qy     299 IGFMPMSLTYQEKYATGSMC----------YTVKDFQEVVKALEDGLISLDKARKMITG 347
           : :   | ||    :           :  |::|  :  | :  |:::    |||
Db     263 YADVTIERFYFEKIVRSELTVWGSWNAISAPFPGKEWQTTIHFLANKQINIE---PMITH 319

Qy     348 KVHLKDGVEKGFKQLIEHKENNVKILVTP 376
           :: | :| |  |::: |  |   |:|  |
Db     320 RLSLAEGPEV-FERIYERNEFFGKVLFFP 347


RESULT 9
AAE05988
ID   AAE05988 standard; Protein; 350 AA.
XX
AC   AAE05988;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Clostridium difficile homologue protein #3.
XX
KW   Five-carbon sugar; aldo-sugar; keto-sugar; sugar alcohol; fermentation;
KW   pentose phosphate pathway; xylitol; D-arabitol; D-arabinose; D-lyxose;
KW   ribitol; D-ribose; D-ribulose; D-xylose; D-xylulose; microbial host;
KW   arabitol phosphate dehydrogenase; APDH; xylitol phosphate dehydrogenase;
KW   XPDH.
XX
OS   Clostridium difficile.
XX
PN   WO200153306-A2.
XX
PD   26-JUL-2001.
XX
PF   22-JAN-2001; 2001WO-FI00051.
```

```
XX
PR   21-JAN-2000; 2000US-0488581.
XX
PA   (XYRO-) XYROFIN OY.
XX
PI   Miasnikov A,  Ojamo H,  Povelainen M,  Gros H,  Toivari M,  Richard P;
PI   Ruohonen L,  Koivuranta K,  Londesborough J,  Aristidou A;
PI   Penttilae M,  Plazanet-Menut C,  Deutscher J;
XX
DR   WPI; 2001-465360/50.
XX
PT   Isolated polynucleotide, used to transform bacterial or yeast hosts
PT   which can then be used in the production of sugars and sugar alcohols,
PT   encodes xylitol phosphate dehydrogenase -
XX
PS   Claim 16; Page 199-200; 205pp; English.
XX
CC   The present invention relates to the methods for manufacturing
CC   five-carbon aldo- and keto-sugars and sugar alcohols by fermentation in
CC   recombinant hosts. The recombinant hosts of the invention have been
CC   engineered to enhance the production of the pentose phosphate pathway
CC   intermediates, or the production of one or more of xylitol, D-arabitol,
CC   D-arabinose, D-lyxose, ribitol, D-ribose, D-ribulose, D-xylose and/or
CC   D-xylulose. Arabitol phosphate dehydrogenase (APDH) is used in a
CC   microbial host cell to produce recombinant arabitol. Xylitol phosphate
CC   dehydrogenase (XPDH) and arabitol phosphate dehydrogenase are used in a
CC   microbial host cell to produce recombinant xylitol. The present sequence
CC   is Clostridium difficile homologue protein which is homologous to
CC   Lactobacillus rhamnosus XPDH protein.
XX
SQ   Sequence   350 AA;

  Query Match             17.7%;  Score 357;  DB 22;  Length 350;
  Best Local Similarity   28.9%;  Pred. No. 8.2e-27;
  Matches  111;  Conservative   59;  Mismatches  144;  Indels   70;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSG----GPVFFP 56
           || : :||  | |  | |  |:|   :|| |||   ||||:|: ::: :|    | :
Db       1 MKSVRFYGIRDTRV-EDVDVPKILEKDDVIIKVKVAGICGSDISKYSKTGPHMVGEIL-- 57

Qy      57 KQGTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVA 116
                       |||||| | :||  | | | ||||||     | :
Db      58 -----------------GHEFSGEVAQVGKEVRSFKIGDRVAVCPAMPCFE--------- 91

Qy     117 QDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEP 176
             |  |: |  : | ::: |     | ||||    | :::|:|| |  :  | :||
Db      92 -----CDECKKGLYSRCNNVAIIGNKELGGCFAEYTKVKERNLIKIPDEISYETAAALEP 146

Qy     177 ISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELG 236
           : :| | : |:  : | | :||| ||||| :|   :   : ||:  :    :   |||||
Db     147 VCIAGHGLFRSEAKVGDTVVVLGTGPIGLFSIQWAKIFGSTKIIAVDVFDEKLDLAKELG 206

Qy     237 AEVFDPSTCDDANAVLKAMVPE------NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVN 290
           |::     |   ||  | :| |       :|     : :| | |    ::      |  :
Db     207 ADI-----C--INAKEKNIVEEIKRLTDGDGADIVIESAGTPLTCGQVLLLAKKGGTVLY 259

Qy     291 VAVWGDHPIGFMPMSLTYQEKYATGSMCYTVK------DFQEVVKALEDGLISLDK---- 340
             |    | | : ::    ||     :  |||        |     |    ||  : |
Db     260 AGV----PYGDVALTREQFEKIVRSEL--TVKGTWFGNSFPFPGKEWSAGLYHMQKGDMN 313

Qy     341 ARKMITGKVHLKDG---VEKGFKQ 361
             |::| :::|::     || :|:
Db     314 VEKLVTHRINLEEAPAYFEKVYKR 337


RESULT 10
ABB60809
ID   ABB60809 standard; Protein; 360 AA.
XX
AC   ABB60809;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 9219.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04912.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 9219; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   360 AA;

  Query Match             17.7%;  Score 355.5;  DB 22;  Length 360;
  Best Local Similarity   33.1%;  Pred. No. 1.2e-26;
  Matches   98;  Conservative   40;  Mismatches  115;  Indels   43;  Gaps    9;

Qy       7 YGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISG 66
           :|  |:|  |  | ||| : ::| : :   ||||:|:   :            :|
Db      11 HGIEDLRL-EQRPIPEIAD-DEVLLAMDSVGICGSDVHYLAHG-------------RIGD 55

Qy      67 YEL--PLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMA 124
           : |  |:  ||| :| | ::|  ||::| |||||:|    |    ||           |
Db      56 FVLTKPMIIGHEAAGVVAKLGKKVTTLKVGDRVAIEPGVPC----RY----------CDH 101

Qy     125 CQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAV 184
           |: |  | || : ||       |    |  :  |   |||| :  : |||:||:||  ||
Db     102 CKQGRYNLCADMVFCATPPYDGNLTRYYKHAADFCFKLPDHVSMEEGALLEPLSVGVHAC 161

Qy     185 ERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE-----V 239
             ||    |   |:|| ||||| |:|| |   | :|: ::    |   ||||||
Db     162 RRAGVGLGSKVLILGAGPIGLVTLLAAQAMGASEILITDLVQQRLDVAKELGATHTLLLQ 221

Qy     240 FDPSTCDDANAVLKAM--VPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
           | |  :    | : |  ||:      : || |   :   :| ||   |: | | :
Db     222 RDQSAEETVKVVHQTMSEVPDK-----SIDCCGAESSARLAIFATRSGGVVVVVGM 272


RESULT 11
AAE05985
ID   AAE05985 standard; Protein; 354 AA.
XX
AC   AAE05985;
```

```
XX
DT   24-SEP-2001  (first entry)
XX
DE   Bacillus halodurans sorbitol dehydrogenase homologue protein #1.
XX
KW   Five-carbon sugar; aldo-sugar; keto-sugar; sugar alcohol; fermentation;
KW   pentose phosphate pathway; xylitol; D-arabitol; D-arabinose; D-lyxose;
KW   ribitol; D-ribose; D-ribulose; D-xylose; D-xylulose; microbial host;
KW   arabitol phosphate dehydrogenase; APDH; xylitol phosphate dehydrogenase;
KW   XPDH; sorbitol dehydrogenase.
XX
OS   Bacillus halodurans.
XX
PN   WO200153306-A2.
XX
PD   26-JUL-2001.
XX
PF   22-JAN-2001; 2001WO-FI00051.
XX
PR   21-JAN-2000; 2000US-0488581.
XX
PA   (XYRO-) XYROFIN OY.
XX
PI   Miasnikov A,  Ojamo H,  Povelainen M,  Gros H,  Toivari M,  Richard P;
PI   Ruohonen L,  Koivuranta K,  Londesborough J,  Aristidou A;
PI   Penttilae M,  Plazanet-Menut C,  Deutscher J;
XX
DR   WPI; 2001-465360/50.
XX
PT   Isolated polynucleotide, used to transform bacterial or yeast hosts
PT   which can then be used in the production of sugars and sugar alcohols,
PT   encodes xylitol phosphate dehydrogenase -
XX
PS   Claim 14; Page 197-198; 205pp; English.
XX
CC   The present invention relates to the methods for manufacturing
CC   five-carbon aldo- and keto-sugars and sugar alcohols by fermentation in
CC   recombinant hosts. The recombinant hosts of the invention have been
CC   engineered to enhance the production of the pentose phosphate pathway
CC   intermediates, or the production of one or more of xylitol, D-arabitol,
CC   D-arabinose, D-lyxose, ribitol, D-ribose, D-ribulose, D-xylose and/or
CC   D-xylulose. Arabitol phosphate dehydrogenase (APDH) is used in a
CC   microbial host cell to produce recombinant arabitol. Xylitol phosphate
CC   dehydrogenase (XPDH) and arabitol phosphate dehydrogenase are used in a
CC   microbial host cell to produce recombinant xylitol. The present sequence
CC   is Bacillus halodurans sorbitol dehydrogenase protein which is
CC   homologous to Lactobacillus rhamnosus XPDH protein.
XX
SQ   Sequence   354 AA;

  Query Match             17.5%;  Score 352.5;  DB 22;  Length 354;
  Best Local Similarity   27.6%;  Pred. No. 2.3e-26;
  Matches  108;  Conservative   56;  Mismatches  176;  Indels   51;  Gaps    9;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || |  ||  |:|: || | | |:: :|: |||   ||||:||  :
Db       1 MKALNLYGIQDLRFEET-PAPSIEHDDDIIIKVKAVGICGSDLSRY---------------- 45

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             |:  |   :  ||||:| ||::|  ||    |||||   |  |              |
Db      46 -KKLGPYVPGMTFGHEFAGEVVKIGRSVTGFSIGDRVAACPTYTC--------------G 90

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           |  || | |  |  ||  |      | :||||     |:: ||: :  |  ||:|| ||
Db      91 QCRYCQLGEPTRCERLSVIG-ARHPGAYAEYVKLPAKHVIPLPNVVNYDEAALIEPASVV 149

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
           |   |  :|| :  ::| | |||  :   :   |  :   :   :   | :|||:|
Db     150 AHGFYRTNIKPGASVAIMGVGSIGLLAVQWAKIFGATTVFAIDIDEQKLNVANQLGADVL 209

Qy     241 DPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIG 300
             |    |: :         |   | : :| | |           |  | : :    |
Db     210 ISSLQRPAHKQILEYT-NGIGVDVAVESAGTPSTSAQVFALPKKGGEVVFLGI----PYA 264

Qy     301 FMPMSLTYQEKYATG----------SMCYTVKDFQEVVKALEDGLISLDKARKMITGKV 349
           : :   | ||               |  : |::   :  :  | :::     ||: ::
Db     265 DVQIERFYFEKIVRNELHVYGSWNALSSPFPGKEWATTIHYMSSGQLNV---APMISYRL 321

Qy     350 HLKDGVEKGFKQLIEHKENNVKILVTPNEVS 380
           |  | |  |:|: : :    |:|  | ::|
Db     322 PLAKGPET-FQQIAKGELKPTKVLFYPEKLS 351


RESULT 12
ABB58718
ID   ABB58718 standard; Protein; 360 AA.
XX
AC   ABB58718;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 2946.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02821.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 2946; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   360 AA;

  Query Match             17.5%;  Score 352.5;  DB 22;  Length 360;
  Best Local Similarity   28.3%;  Pred. No. 2.4e-26;
  Matches  106;  Conservative   63;  Mismatches  161;  Indels   45;  Gaps   11;

Qy       7 YGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISG 66
           :|  |:|  |  | ||| : ::| : :   ||||:|:    :          :|
Db      11 HGIEDMRL-EQRPIPEIAD-DEVLLAMDSVGICGSDVHYLAHG----------RIGD 55

Qy      67 YEL--PLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMA 124
           : |  |:  ||| :| | ::|  ||::| |||||:|    |               |
Db      56 FVLTKPMIIGHESAGVVAKLGKKVTTLKVGDRVAIEPGVPCRK----------CDH 101
```

```
Qy    125 CQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAV 184
          |: |  | |  : ||      |    |  :  |   |||| :  : |||:||:||  ||
Db    102 CKQGKYNLCPGMVFCATPPYDGNLTRYYKHAADFCFKLPDHVTMEEGALLEPLSVGVHAC 161

Qy    185 ERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE-----V 239
          :||    |   |:|| ||||| |::| |   | :|: ::   |   ||||||
Db    162 KRAEVTLGSKVLILGAGPIGLVTLMAAQAMGASEILITDLVQQRLDVAKELGATHTLLLK 221

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           | :  : |  | | |  : :    : || |   :   :| ||   || | | : |   |
Db    222 RDQTAEETAVLVQKTMGGQPD---KSIDCCGAESSARLAIFATRSGGIVVVVGM-GAAEI 277

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
            ::   :|    |   |   |:   :  :  | :::   ::::|    :|:   | |
Db    278 KLPLINALAREVDIRGVFRY-CNDYAAALALVASGKVNV---KRLVTHHFDIKE-TAKAF 332

Qy    360 KQLIEHKENNVKILV 374
          :   :    :|:::
Db    333 ETSRKGLGGAIKVMI 347
```

RESULT 13
AAB47473
ID AAB47473 standard; Protein; 316 AA.
XX
AC AAB47473;
XX
DT 15-NOV-2001 (first entry)
XX
DE G. max sorbitol dehydrogenase.
XX
KW Aldehyde reductase; clustal alignment; plant specific promoter;
KW carbohydrate metabolism; sorbitol biosynthesis; carbon partitioning;
KW NADPH-dependant mannose 6-phosphate dehydrogenase;
KW sorbitol dehydrogenase.
XX
OS Glycine max.
XX
PN US6274379-B1.
XX
PD 14-AUG-2001.
XX
PF 02-JUL-1999; 99US-0347803.
XX
PR 15-JUL-1998; 98US-0092952.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Famodu LO, Kinney AJ, Orozco EM;
XX
DR WPI; 2001-556714/62.
DR N-PSDB; AAH43316.
XX
PT Nucleic acid sequence encoding a polypeptide having aldehyde reductase
PT activity, useful for transforming plants with altered levels of
PT expression of sorbitol biosynthetic enzymes -
XX
PS Example 5; Column 51-52; 33pp; English.
XX
CC The sequences given in AAH43311-17 encode enzymes which were used
CC for comparison in the isolation of proteins having aldehyde reductase
CC activity. The isolated aldehyde reductase sequences were identified
CC based on Clustal alignment method. The aldehyde reductase cDNA,
CC operatively linked to a plant specific promoter, is useful for
CC altering the level of expression of aldehyde reductase in a
CC transformed host cell. The aldehyde reductase is also useful in
CC studies facilitating better understanding of carbohydrate metabolism
CC and function in plants, for providing genetic tools for the manipulation
CC of the sorbitol biosynthetic pathway, and for providing means to control
CC carbon partitioning in plant cells.
XX
SQ Sequence 316 AA;

Query Match 17.3%; Score 349; DB 22; Length 316;
Best Local Similarity 33.3%; Pred. No. 4.4e-26;
Matches 88; Conservative 36; Mismatches 112; Indels 28; Gaps 4;

```
Qy     26 PNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVVEVG 85
          |:||::::   ||||:|:          |  |:           |:  ||| :| : |||
Db     41 PHDVRVRMKAVGICGSDVHYLKTLRCAHFIVKE----------PMVIGHECAGIIEEVG 89

Qy     86 SGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLSFCGLGGAS 145
          | | .|: ||||||:|    |                |  |: |  | |  : |
Db     90 SQVKSLVPGDRVAIEPGISCWH-------------CNHCKHGRYNLCDDMKFFATPPVH 135

Qy    146 GGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTALVLGGGPIGL 205
          |  |  :|:  |   ||||::  : ||: ||:||  ||  ||   |    |::| |||||
Db    136 GSLANQIVHPADLCFKLPDNVSLEEGAMCEPLSVGVHACRRANIGPETNVLIMGAGPIGL 195

Qy    206 ATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPST--CDDANAVLKAMVPENEGF 262
           |:|| :   | | |  :    |   || ||| ::   ||   | |  |::      |
Db    196 VTMLAARAFGAPKTVIVDVDDHRLSVAKSLGADDIIKVSTNIKDVAEEVVQIQKVMGAGI 255

Qy    263 HAAFDCSGVPQTFTTSIVATGPSG 286
             |||:|  :| : :: || | |
Db    256 DVTFDCAGFDKTMSPALSATQPGG 279
```

RESULT 14
ABB49510
ID ABB49510 standard; Protein; 343 AA.
XX
AC ABB49510;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #2214.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 6; SEQ ID No 2215; 192pp; French.
XX
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins

```
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   343 AA;

  Query Match             17.3%;  Score 348;  DB 23;  Length 343;
  Best Local Similarity   31.6%;  Pred. No. 6.3e-26;
  Matches  115;  Conservative   51;  Mismatches  150;  Indels   48;  Gaps   13;

Qy      18 VPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEF 77
           | ||::   : |||||:: ||||:|:  |              | :      |:  ||||
Db      19 VEEPQVYG-DKVKIKVAFTGICGSDIHTF-------------KGEYKNPTTPVTLGHEF 63

Qy      78 SGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLS 137
           || |||||  |||:| ||||  |              |  :  | |: |:    | |::
Db      64 SGVVVEVGPDVTSIKVGDRVTSE------------TTFETCGECIYCKEHDYNLCSNRR 110

Qy     138 FCGLG-GASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWH-AVERARFQPGQTA 195
            |:|  |:| |||:|:  |:    | : |  :  || ||::   | |:|:   :|  |
Db     111 --GIGTQANGSFAEFVLSREESCHVLDERISLEAAALTEPLACCVHSALEKTTIRPDDTV 168

Qy     196 LVLGGGPIG--LATILALQGH---HAGKIVCSEPALIRRQFAKELGAEVFDPSTCDDANA 250
           || | ||||  || ::  ||     ||    |:    | : ||||| :    :  :|
Db     169 LVFGPGPIGLLLAQVVKAQGATVIMAGITKDSD----RLRLAKELGMDRIVDTLKEDLAE 224

Qy     251 VLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTYQE 310
           |:  |     |   |||||         :  |   |  | | :: :        |:  :|
Db     225 VVLGMT-GGYGAERVFDCSGAVPAVNQGLPLTKKKGDFVQVGLFAEKKNAIDEESIIQRE 283

Qy     311 KYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKENNV 370
           ||      :  :  | :| |  |   |||| ||:  |   : |: ::    | :
Db     284 IAYIGSRSQKPSSWILALDLLANGKIDTD---KMIT-KVYGLDDWREAFEAVM--AGNEI 337

Qy     371 KILV 374
           |:||
Db     338 KVLV 341


RESULT 15
AAM47254
ID   AAM47254 standard; Protein; 368 AA.
XX
AC   AAM47254;
XX
DT   20-FEB-2002  (first entry)
XX
DE   Arxula adeninivorans membrane protein.
XX
KW   Yeast; protein production; AXOH; insulin; phytase; hirudin;
KW   human growth hormone; lactoferrin; granulocyte colony-stimulating factor;
KW   membrane protein.
XX
OS   Arxula adeninivorans.
XX
FH   Key             Location/Qualifiers
FT   Region          80..94
FT                   /note= "ZnADH signature"
FT   Region          188..214
FT                   /note= "transmembrane region"
XX
PN   WO200185925-A2.
XX
PD   15-NOV-2001.
XX
PF   08-MAY-2001; 2001WO-EP05225.
XX
PR   08-MAY-2000; 2000DE-1022334.
XX
PA   (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
PA   (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
XX
PI   Kunze G,  Walter I,  Roesel H,  Boeer E,  Stoltenburg R,  Wartmann T;
PI   Duc BM;
XX
DR   WPI; 2002-055593/07.
DR   N-PSDB; ABA03453, ABA03454.
XX
PT   Producing proteins in Arxula yeast, useful e.g. for preparing insulin,
PT   by expression under control of specific inducible or constitutive
PT   Arxula promoters -
XX
PS   Example 15; Fig 31; 125pp; German.
XX
CC   The present invention relates to a method of producing proteins in yeast
CC   from the genus Arxula, involving cloning a nucleic acid into an
CC   expression vector under the control of an inducible promoter from Arxula,
CC   transforming Arxula cells with the vector, culturing the cells, inducing
CC   the promoter and harvesting the desired protein. This can be used to
CC   produce proteins including insulin, phytase, human growth hormone,
CC   hirudin, lactoferrin and granulocyte colony-stimulating factor. The
CC   present sequence is an Arxula adeninivorans membrane protein coded for by
CC   the AXOH gene.
XX
SQ   Sequence   368 AA;

  Query Match             17.2%;  Score 346;  DB 23;  Length 368;
  Best Local Similarity   31.7%;  Pred. No. 1.1e-25;
  Matches  107;  Conservative   40;  Mismatches  153;  Indels   38;  Gaps    8;

Qy      16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGY--ELPLCP 73
           |  | | | :| ||||:|   ||||:|          | | : |   :|. |  | |:
Db      33 EERPVPVITDPRDVKIQVKKTGICGSD----------VHFWQHG---RIGDYVVEKPMVL 79

Qy      74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
           ||| || ||||||  |||:| |||||:|          ||               |  : |
Db      80 GHESSGVVVEVGSEVTSLKVGDRVAMEPGVPDRRSKEYK-------------MGRYHLC 125

Qy     134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
             : |       |   :|    ||  ||||:::  : ||||||:||| |         ||
Db     126 PHVRFAACPPTDGTLCKYYTLPEDFCVKLPENVDFEEGALVEPLSVAVHTARLLGIYPGS 185

Qy     194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPSTCDD----A 248
             :| | ||||   |   :   | |   :    : : ||| ||   : |   |      |
Db     186 KVVVFGAGPIGQLCIGVCKAFGASIIGAVDLFEQKLETAKEFGASHTYVPQKGDSHDETA 245

Qy     249 NAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTY 308
           : :|: ::|  :      | ||  |:    |      |    ||:     | |    :
Db     246 HKILE-LLPNKQAPDVVIDASGAEQSINAGIELLERGGTFGQVAMGRTDYIQFAVSRMAM 304

Qy     309 QEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMIT 346
           :|    |   ||  |::   : : || | :   :|::|
Db     305 KEIRFQGVFRYTYGDYELATQLIGDGKIPV---KKLVT 339


Search completed: May  2, 2003, 12:24:31
Job time : 941 secs
```

OM protein - protein search, using sw model

Run on: May 2, 2003, 12:05:52 ; Search time 622 Seconds
(without alignments)
58.732 Million cell updates/sec

Title: US-10-020-674-2
Perfect score: 2013
Sequence: 1 MKGLLYYGTNDIRYSETVPE.....QLIEHKENNVKILVTPNEVS 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	923.5	45.9	382	2	S51962	FUN49 protein - ye
2	912	45.3	417	2	S51961	FUN50 protein - ye
3	571	28.4	348	2	E84143	L-iditol 2-dehydro
4	567.5	28.2	363	2	C83128	2,3-butanediol deh
5	551.5	27.4	346	1	H69789	probable alcohol d
6	538.5	26.8	370	2	A86737	(R,R)-butanediol d
7	537	26.7	354	2	E81178	probable alcohol d
8	534	26.5	354	2	H81925	probable alcohol d
9	455.5	22.6	352	2	B83133	probable alcohol d
10	422.5	21.0	357	2	AH3139	zinc-binding dehyd
11	422.5	21.0	397	2	D98148	(R,R)-butanediol d
12	408.5	20.3	353	1	A45052	L-iditol 2-dehydro
13	386.5	19.2	347	2	T23889	hypothetical prote
14	370.5	18.4	348	1	S32484	L-iditol 2-dehydro
15	370	18.4	395	2	B72381	alcohol dehydrogen
16	369	18.3	339	1	F69852	L-iditol 2-dehydro
17	368.5	18.3	347	2	T23890	hypothetical prote
18	365.5	18.2	357	1	S67811	L-iditol 2-dehydro
19	364.5	18.1	348	2	AB1496	polyol (sorbitol)
20	362	18.0	360	1	S35981	L-iditol 2-dehydro
21	360.5	17.9	357	1	S55941	L-iditol 2-dehydro
22	358.5	17.8	348	2	AC1138	polyol (sorbitol)
23	352.5	17.5	354	2	C83673	sorbitol dehydroge
24	348	17.3	343	2	AF1407	polyol dehydrogena
25	348	17.3	343	2	AF1783	polyol dehydrogena
26	346	17.2	347	2	AH2749	sorbitol dehydroge
27	346	17.2	371	2	G97530	probable zinc-bind
28	345	17.1	358	2	E90939	probable oxidoredu
29	345	17.1	358	2	A85788	probable oxidoredu
30	343	17.0	343	2	E95295	IdnD L-idonate 5-d
31	342	17.0	358	1	H64937	probable alcohol d
32	342	17.0	363	1	S13529	D-xylulose reducta
33	339	16.8	343	2	F98227	L-idonate 5-dehydr
34	339	16.8	343	2	AB3059	zinc-binding dehyd
35	338	16.8	371	2	AB0407	probable zinc-bind
36	337.5	16.8	343	2	E83673	sorbitol dehydroge
37	337.5	16.8	389	2	AF1918	alcohol dehydrogen
38	337.5	16.8	399	1	S16132	L-iditol 2-dehydro
39	337	16.7	347	1	F64937	probable L-iditol
40	336.5	16.7	354	1	S10065	L-iditol 2-dehydro
41	336	16.7	357	1	A54674	L-iditol 2-dehydro
42	333.5	16.6	356	1	S64902	probable sugar red
43	331.5	16.5	348	2	G75049	L-threonine 3-dehy
44	330	16.4	347	2	C90939	probable oxidoredu
45	330	16.4	347	2	G85787	probable oxidoredu

ALIGNMENTS

RESULT 1
S51962
FUN49 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YAL060w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S51962
R;Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.;
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A;Reference number: S51956
A;Accession: S51962
A;Molecule type: DNA
A;Residues: 1-382 <BUS>
A;Cross-references: EMBL:U12980; NID:g1326053; PID:g595526; GSPDB:GN00001; MIPS:YAL0
C;Genetics:
A;Gene: SGD:BDH1; FUN49; MIPS:YAL060w
A;Cross-references: SGD:S0000056
A;Map position: 1L
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
F;24-363/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 45.9%; Score 923.5; DB 2; Length 382;
Best Local Similarity 48.7%; Pred. No. 2.9e-63;
Matches 184; Conservative 65; Mismatches 126; Indels 3; Gaps 2;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           |: | |:   || ::  :| |||:  ::| | ||:||||||:|| |  |  ||:| || |
Db       1 MRALAYFKKGDIHFTNDIPRPEIQTDDEVIIDVSWCGICGSDLHE--YLDGPIFMPKDGE 58

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             |:|   |||  ||| || | :||  || || || | |:| | |:|   :  :   :
Db      59 CHKLSNAALPLAMGHEMSGIVSKVGPKVTKVKVGDHVVVDAASSCADLHCWPHSKFYNSK 118

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           | ||| || | |    | |||  |||||| ||  : |:: :|  || |: ||||:||
Db     119 PCDACQRGSENLCTHAGFVGLGVISGGFAEQVVVSQHHIIPVPKEIPLDVAALVEPLSVT 178

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
           ||||: : |: | :|||||| ||||| ||| |:|  | ||| || |  | : ||:|| |||
Db     179 WHAVKISGFKKGSSALVLGAGPIGLCTILVLKGMGASKIVVSEIAERRIEMAKKLGVEVF 238

Qy     241 DPSTCDDAN-AVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           :||     :  :|: :   ::||  ::||||:  || ||: |    | | |:|||  |:
Db     239 NPSKHGHKSIEILRGLTKSHDGFDYSYDCSGIQVTFETSLKALTFKGTATNIAVWGPKPV 298

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
           | || :| |||  |||: | |: |:|||:|: :| |:::  :::|||| ::|| ||||
Db     299 PFQPMDVTLQEKVMTGSIGYVVEAFEEVVRAIHNGDIAMEDCKQLITGKQRIEDGWEKGF 358

Qy     360 KQLIEHKENNVKILVTPN 377
           ::|::|||:||||||:|||
```

```
Db    359 QELMDHKESNVKILLTPN 376


RESULT 2
S51961
FUN50 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YAL061w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S51961
R;Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A;Reference number: S51956
A;Accession: S51961
A;Molecule type: DNA
A;Residues: 1-417 <BUS>
A;Cross-references: EMBL:U12980; NID:g1326053; PID:g595525; GSPDB:GN00001; MIPS:YAL061w
C;Genetics:
A;Gene: FUN50; MIPS:YAL061w
A;Cross-references: SGD:S0000057
A;Map position: 1L
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
F;24-364/Domain: long-chain alcohol dehydrogenase homology <LADH>

  Query Match             45.3%;  Score 912;  DB 2;  Length 417;
  Best Local Similarity   45.9%;  Pred. No. 2.5e-62;
  Matches  174;  Conservative   71;  Mismatches  130;  Indels    4;  Gaps    2;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          |: | |:|  :||::  : || |  |::: | : :||||||| |  |: ||:|||: |
Db      1 MRALAYFGKGNIRFTNHLKEPHIVAPDELVIDIEWCGICGTDLHE--YTDGPIFFPEDGH 58

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
            :||   ||   ||| :|||:||| || ::| ||:| || |  | || |:  :   |
Db     59 THEISHNPLPQAMGHEMAGTVLEVGPGVKNLKVGDKVVVEPTGTCRDRYRWPLSPNVDKE 118

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           | ||: |  | |: |  || |  |||||| ||  | |  |:|| :| |: ||::|::|
Db    119 WCAACKKGYYNICSYLGLCGAGVQSGGFAERVVMNESHCYKVPDFVPLDVAALIQPLAVC 178

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
          |||:    |: | |||::| ||||| |||||       || |||| :||: |:::|| |:
Db    179 WHAIRVCEFKAGSTALIIGAGPIGLGTILALNAAGCKDIVVSEPAKVRRELAEKMGARVY 238

Qy    241 DPS--TCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHP 298
          ||:     ::   |:::    :||   |||||:  |   :|      | |||:|:|| |
Db    239 DPTAHAAKESIDYLRSIADGGDGFDYTFDCSGLEVTLNAAIQCLTFRGTAVNLAMWGHHK 298

Qy    299 IGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358
          | | || :|  |:  |||||||  ||: |::|||:| | :|:|| ||||:|:::||::
Db    299 IQFSPMDITLHERKYTGSMCYTHHDFEAVIEALEEGRIDIDRARHMITGRVNIEDGLDGA 358

Qy    359 FKQLIEHKENNVKILVTPN 377
            :||  ||: :||::|||
Db    359 IMKLINEKESTIKIILTPN 377


RESULT 3
E84143
L-iditol 2-dehydrogenase BH3949 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: E84143
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: E84143
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <STO>
A;Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BAB07668.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3949
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

  Query Match             28.4%;  Score 571;  DB 2;  Length 348;
  Best Local Similarity   37.7%;  Pred. No. 3e-36;
  Matches  141;  Conservative   65;  Mismatches  134;  Indels   34;  Gaps   13;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          |: | ::|  |:|  :      :|    ||||| :|||||:|| |  |: ||:| | | :
Db      1 MQALRWHGVKDVRVDDIAEPATLK--GKVKIKVKWCGICGSDLHE--YAAGPIFIP-QNS 55

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
            ::  : |:  |||||| ||||| |||  : |||| ||  .  |                |
Db     56 PHSLTEEKAPIVMGHEFSGQVVEVGDGVTKCEEGDRVVVEPIFAC-------------G 101

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMV-KLPDSIPDDIGALVEPISV 179
           | ||: |  | |  | | || |  |||:|||   ::||| |:||::  : |||||| :|
Db    102 TCTACKQGKYNLCEQLGFLGLAGGGGGFSEYVTV-DEHMVHKIPDTVSFEQGALVEPAAV 160

Qy    180 AWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEV 239
          | :|| :::|  |  |:| | ||||| | ||:  | ||   | :  ||| |::|||
Db    161 ALYAVRQSQFNVGDQAVVFGTGPIGLLVIEALKASGASKIYAVELSTERRQRAEQLGAIA 220

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           :|:   : | | :     : |   :|: :|||  | :| :|  :|  : |::: :
Db    221 LNPA---EVNVVEEIQRLTDGGADVSFEVTGVPIVLTQAIESTKLNGQTMIVSIF-EKEA 276

Qy    300 GFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358
          | |  :  :|:   | :.|  :| | |:  :  |  |   |  ::| :: | | ||:|
Db    277 SFQPNLVVMRERNIAGIIGY--RDIFPAVISLMNQGYFS---AETLVTKRIGLHDIVEQG 331

Qy    359 FKQLIEHKEN-NVK 371
          |: |:  | | |:|
Db    332 FEALM--KTNCNIK 343


RESULT 4
C83128
2,3-butanediol dehydrogenase PA4153 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: C83128
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83128
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-363 <STO>
A;Cross-references: GB:AE004831; GB:AE004091; NID:g9950347; PIDN:AAG07540.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4153
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

  Query Match             28.2%;  Score 567.5;  DB 2;  Length 363;
  Best Local Similarity   36.1%;  Pred. No. 5.8e-36;
  Matches  136;  Conservative   73;  Mismatches  141;  Indels   27;  Gaps   11;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
            |:  :::| :|||  | || |    |  |:|:| :|||||:|| |  |  |||| | :
Db     11 MRAAVWHGRHDIRV-EDVPLPAEPPPGWVQIRVHWCGICGSDLHE--YLAGPVFIPVE-A 66

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
              ::| :     |||||| :| :|:|||    |  || :|  ||           |
Db     67 PHPLTGLKDQCILGHEFSGEIVRLGNGVTGFAVGQAVAADACQHC-------------G 112

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
```

```
Db    113 TCYYCRHGLYNICENLAFTGLMN-NGAFAEYVNVPANLLYALPAGFPSEAGALIEPLAVG 171

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239

Db    172 MHAVKKAGSLLGQNVVVVGAGTIGLSTIMCARAAGAAQVIALEMSSARKAKALEVGASQV 231

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299

Db    232 LDPSRC-DALGEIRALT-GGLGADVSFECIGNKHTAKLAIDAIRKAGKCVLVGIF-EEPS 288

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359

Db    289 EFNFFELVSTEKQLLGALAYN-GEFADVIAFIADGRLDI---APLVTGRIGLEEIVERGF 344

Qy    360 KQLIEHKENNVKILVTP 376

Db    345 EELVNNKEHNVKIIVSP 361


RESULT 5
H69789
probable alcohol dehydrogenase (EC 1.1.1.-) ydjL [similarity] - Bacillus subtilis
N;Alternate names: L-iditol 2-dehydrogenase homolog ydjL; sorbitol dehydrogenase homolog
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Sep-2000
C;Accession: H69789
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: H69789
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-346 <KUN>
A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12443.1; PID:g2632937
A;Experimental source: strain 168
C;Genetics:
A;Gene: ydjL
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: metalloprotein; NAD; oxidoreductase; zinc
F;22-334/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;171-200/Region: beta-alpha-beta NAD nucleotide-binding fold
F;37,70,152/Binding site: zinc, catalytic (Cys, His, Glu) #status predicted

  Query Match             27.4%;  Score 551.5;  DB 1;  Length 346;
  Best Local Similarity   38.4%;  Pred. No. 9.2e-35;
  Matches  147;  Conservative   55;  Mismatches  138;  Indels   43;  Gaps   13;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Db      1 MKAARWHNQKDIRI-EHIEEPKTE-PGKVKIKVKWCGICGSDLHE--YLGGPIFIP---- 52

Qy     61 KDK---ISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEA--TSHCSDRSRYKDTV 115

Db     53 VDKPHPLTNETAPVTMGHEFSGEVVEVGEGVENYKVGDRVVVEPIFATH---------- 101

Qy    116 AQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVE 175

Db    102 ---------GHQGAYNLDEQMGFLGLAGGGGGFSEYVSVDEELLFKLPDELSYEQGALVE 152

Qy    176 PISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKEL 235

Db    153 PSAVALYAVRSSKLKAGDKAAVFGCGPIGLLVIEALKAAGATDIYAVELSPERQQKAEEL 212

Qy    236 GAEVFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWG 295

Db    213 GAIIVDPSKTDDVVAEIAERT--GGGVDVAFEVTGVPVVLRQAIQSTTIAGETVIVSIW- 269

Qy    296 DHPIGFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDG 354

Db    270 EKGAEIHPNDIVIKERTVKGIIGY--RDIFPAVLSLMKEGYFSAD---KLVTKKIVLDDL 324

Qy    355 VEKGFKQLIEHKENNVKILVTPN 377

Db    325 IEEGFGALIKEK-SQVKILVRPN 346


RESULT 6
A86737
(R,R)-butanediol dehydrogenase (EC 1.1.1.4) [imported] - Lactococcus lactis subsp. l
N;Alternate names: 2,3-butanediol dehydrogenase
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: A86737
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; E
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lacti
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: A86737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <STO>
A;Cross-references: GB:AE005176; PID:g12723828; PIDN:AAK04995.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: butB
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: oxidoreductase

  Query Match             26.8%;  Score 538.5;  DB 2;  Length 370;
  Best Local Similarity   37.3%;  Pred. No. 9.9e-34;
  Matches  144;  Conservative   53;  Mismatches  144;  Indels   45;  Gaps   14;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Db     15 MRAARFYDRGDIRIDE-INEPIVK-AGQVGIDVAWCGICGTDLHEFL--DGPIFCPSAEH 70

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--ATSHCSDRSRYKDTVAQD 118

Db     71 PNPITGEVPPVTLGHEMSGVVNFIGEGVSGLKVGDHVVVEPYIVPEGTDTS--------- 121

Qy    119 LGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPIS 178

Db    122 -------ETGHYNLSEGSNFIGLGGNGGGLAEKISVDERWVHKIPDNLPLDEAALIEPLS 174

Qy    179 VAWHAVERARFQPGQTALVLGGGPIGL--ATILALQGHHAGKIVCSEPALIRRQFAKELG 236

Db    175 VGYHAVERANLSEKSTVLVVGAGPIGLLTAAVAKAQGH---TVIISEPSGLRRKKAQE-- 229

Qy    237 AEV----FDPSTCDDANAVLKAMVPE--NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVN 290

Db    230 AQVADYFFNPIEDD-----IQAKVHEINEKGVDAAFECTSVQPGFDACLDAIRMGGTVVI 284

Qy    291 VAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVH 350

Db    285 VAIWG-KPASVDMAKLVIKEANLLGTIAYN-NTHPKTIDLVSTGKIKLD---QFITAKIG 339

Qy    351 LKDGVEKGFKQLIEHKENNVKILVTP 376

Db    340 LDDLIDKGFDTLIHHNETAVKILVSP 365


RESULT 7
```

```
E81178
probable alcohol dehydrogenase (EC 1.1.1.-) NMB0604 [similarity] - Neisseria meningitidi
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: E81178
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: E81178
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-354 <TET>
A;Cross-references: GB:AE002416; GB:AE002098; NID:g7225832; PIDN:AAF41031.1; PID:g722583
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0604
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: oxidoreductase

  Query Match             26.7%;  Score 537;  DB 2;  Length 354;
  Best Local Similarity   36.0%;  Pred. No. 1.2e-33;
  Matches  140;  Conservative   54;  Mismatches  147;  Indels   48;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||    :|   |||  | :||| :  |  | | |::||||||||| ||    ||:| |  |
Db       1 MKAARFYDKGDIRI-EDIPEPTVA-PGTVGINVAWCGICGTDLHEF--MEGPIFIPPCGH 56

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--------ATSHCSDRSRYK 112
             |||   |:  |||||| |  || ||  :| |  | ||          |   |:    |
Db      57 PHPISGESAPVTMGHEFSGVVYAVGEGVDDIKVGQHVVVEPYIIRDDVPTGEGSNYHLSK 116

Qy     113 DTVAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGA 172
           |                       ::| ||||  || :| :       :  : | |  |  |
Db     117 D----------------------MNFIGLGGCGGGLSEKIAVKRRWVHPISDKIPLDQAA 154

Qy     173 LVEPISVAWHAVERARFQPGQTALVLGGGPIG--LATILALQGHHAGKIVCSEPALIRRQ 230
           |:||:||  ||  |:  : |  ||| | ||||  || :|  :|    |:: :| :  |:
Db     155 LIEPLSVGHHAYVRSGAKEGDVALVGGAGPIGLLLAAVLKAKGI---KVIITELSKARKD 211

Qy     231 FAKELGAE--VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIA 288
            |:| |    : |||  |    | |  :   ||   ||:|: | :   | : |  |:
Db     212 KARESGVADYILDPSEVDVVAEVKK--LTNGEGVDVAFECTSVNKVLDTLVEACKPAANL 269

Qy     289 VNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGK 348
           | |::|  ||      |:  :|    |:: |   |  | :| :|:| |:|:     || :
Db     270 VIVSIW-SHPATINVHSVVMKELDVRGTIAY-CNDHAETIKLVEEGKINLE---PFITQR 324

Qy     349 VHLKDGVEKGFKQLIEHKENNVKILVTPN 377
           : | : | |||::|| : |: |||:|:||
Db     325 IKLDELVSKGFERLIHNNESAVKIIVSPN 353


RESULT 8
H81925
probable alcohol dehydrogenase (EC 1.1.1.-) NMA0808 [similarity] - Neisseria meningitidi
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: H81925
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H81925
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-354 <PAR>
A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84091.1; PID:g737952
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0808
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: oxidoreductase

  Query Match             26.5%;  Score 534;  DB 2;  Length 354;
  Best Local Similarity   35.7%;  Pred. No. 2.1e-33;
  Matches  139;  Conservative   55;  Mismatches  147;  Indels   48;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||    :|   |||  | :||| :  |  | | |::||||||||| ||    ||:| |  |
Db       1 MKAARFYDKGDIRI-EDIPEPTVA-PGTVGINVAWCGICGTDLHEF--MEGPIFIPPCGH 56

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE--------ATSHCSDRSRYK 112
             |||   |:  |||||| |  || ||  :| |  | ||          |   |:    |
Db      57 PHPISGESAPVTMGHEFSGVVYAVGEGVDDIKVGQHVVVEPYIIHDDVPTGEGSNYHLSK 116

Qy     113 DTVAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGA 172
           |                       ::| ||||  || :| :       :  : | || |  |
Db     117 D----------------------MNFIGLGGCGGGLSEKIAVKRRWVHPISDKIPLDQAA 154

Qy     173 LVEPISVAWHAVERARFQPGQTALVLGGGPIG--LATILALQGHHAGKIVCSEPALIRRQ 230
           |:||:||  ||  |:  : |  ||| | ||||  || :|  :|    |:: :| :  |:
Db     155 LIEPLSVGHHAYVRSGAKAGDVALVGGAGPIGLLLAAVLKAKGI---KVIITELSKARKD 211

Qy     231 FAKELGAE--VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIA 288
            |:| |    : |||  |    | |  :   ||   ||:|: | :   | : |  |:
Db     212 KARESGVADYILDPSEVDVVEEVKK--LTNGEGVDVAFECTSVNKVLDTLVEACKPAANL 269

Qy     289 VNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGK 348
           | |::|  ||      |:  :|    |:: |   |  | :| :|:| |:|:     || :
Db     270 VIVSIW-SHPATVNVHSVVMKELDVRGTIAY-CNDHAETIKLVEEGKINLE---PFITQR 324

Qy     349 VHLKDGVEKGFKQLIEHKENNVKILVTPN 377
           : | : | |||::|| : |: |||:|:||
Db     325 IKLDELVSEGFERLIHNNESAVKIIVSPN 353


RESULT 9
B83133
probable alcohol dehydrogenase (Zn-dependent) PA4097 [imported] - Pseudomonas aerugin
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83133
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.,
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83133
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-352 <STO>
A;Cross-references: GB:AE004826; GB:AE004091; NID:g9950296; PIDN:AAG07484.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4097
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

  Query Match             22.6%;  Score 455.5;  DB 2;  Length 352;
  Best Local Similarity   33.5%;  Pred. No. 2.2e-27;
  Matches  127;  Conservative   55;  Mismatches  162;  Indels   35;  Gaps   12;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || | ::   |:| ||   | :   | :|:::|:||||||:|| | |  ||   | |
Db       1 MKALRWHAARDLRLSEL--ERQAPRPGEVELEVAYCGICGSDLHE--YQSGPHSIP-QAE 55

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             :||   ||  |||| | |  :| ||   : |||||||    |                |
Db      56 AHPLSGCRAPLTLGHEFCGVVAALGPGVEGPRIGDRVAVEPEYRC--------------G 101

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
```

```
Db    102 ECRYCREGRYNLCESMGFIGLMG-DGGFAERARVPAYMLHRLPDAVGFRQAAVLEPAAVA 160

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239

Db    161 LHALRRSSLAPGQRCAVFGLGPIGLLLVMLARLRGIEDIAAVDVSPERLALAGEFGASRA 220

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299

Db    221 LDARDGDTA-----ARLREGGALDCAFEAAGSQASLDAALASLRKGGELVLVSLMGE--V 273

Qy    300 GFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358

Db    274 RLDAFDLVNRELRLLGSVGY--RDAYPELIALLADGRLDLARA---VTRSVPLEQAVEHG 328

Qy    359 FKQLIEHKENNVKILVTPN 377

Db    329 FEALLRDK-SQLKVLVNPN 346


RESULT 10
AH3139
zinc-binding dehydrogenase Atu4740 [imported] - Agrobacterium tumefaciens (strain C58, D
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AH3139
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AH3139
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45534.1; PID:g17743246; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4740
A;Map position: linear chromosome
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

  Query Match           21.0%;  Score 422.5;  DB 2;  Length 357;
  Best Local Similarity  28.5%;  Pred. No. 7.6e-25;
  Matches  109;  Conservative   75;  Mismatches  160;  Indels   39;  Gaps   10;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Db      1 MRALRFHAAKDLRI-EDIAEPKRPGPGQVLVRNRFVGICGTDLHEYSY--GPIFIPTE-- 55

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE------ATSHCSDRSRYKDT 114

Db     56 PHPFTGAHGPQILGHEFGGVVEAIGDGVTSVNVGDRVSIQPLIMPRSGDYFADRGLF--- 112

Qy    115 VAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALV 174

Db    113 ---------------HLSTQLALVGLSWDGGGMAEAALVNEYNVQKIPDEMTDEEAALV 156

Qy    175 EPISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKE 234

Db    157 EPSAVAVYACDRGGVTAGNSVLVTGAGPIGMLTLLAARAAGATQLFVSDLNDARLELARN 216

Qy    235 LGAEV--FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVA 292

Db    217 VIRDVITINPKR-DNVGDVVRSATEGKVGCDVAIECVGNEHALKACVDAVRKQGVVVQTG 275

Qy    293 VW-GDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHL 351

Db    276 LHPHENPIDW--FQVTFRDLEIKGSWAYPTHYWPRVIRLIASGLL---PATKIVTKRITL 330

Qy    352 KDGVEKGFKQLIEHKENNVKILV 374

Db    331 DTAVTEGFDALLDPAGTHLKILI 353


RESULT 11
D98148
(R,R)-butanediol dehydrogenase (EC 1.1.1.4) [imported] - Agrobacterium tumefaciens (
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: D98148
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; PMID:11743194
A;Accession: D98148
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-397 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88710.1; PID:g15158444; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_281
A;Map position: linear chromosome
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: oxidoreductase

  Query Match           21.0%;  Score 422.5;  DB 2;  Length 397;
  Best Local Similarity  28.5%;  Pred. No. 8.7e-25;
  Matches  109;  Conservative   75;  Mismatches  160;  Indels   39;  Gaps   10;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Db     41 MRALRFHAAKDLRI-EDIAEPKRPGPGQVLVRNRFVGICGTDLHEYSY--GPIFIPTE-- 95

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVE------ATSHCSDRSRYKDT 114

Db     96 PHPFTGAHGPQILGHEFGGVVEAIGDGVTSVNVGDRVSIQPLIMPRSGDYFADRGLF--- 152

Qy    115 VAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALV 174

Db    153 ---------------HLSTQLALVGLSWDGGGMAEAALVNEYNVQKIPDEMTDEEAALV 196

Qy    175 EPISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKE 234

Db    197 EPSAVAVYACDRGGVTAGNSVLVTGAGPIGMLTLLAARAAGATQLFVSDLNDARLELARN 256

Qy    235 LGAEV--FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVA 292

Db    257 VIRDVITINPKR-DNVGDVVRSATEGKVGCDVAIECVGNEHALKACVDAVRKQGVVVQTG 315

Qy    293 VW-GDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHL 351

Db    316 LHPHENPIDW--FQVTFRDLEIKGSWAYPTHYWPRVIRLIASGLL---PATKIVTKRITL 370

Qy    352 KDGVEKGFKQLIEHKENNVKILV 374

Db    371 DTAVTEGFDALLDPAGTHLKILI 393


RESULT 12
A45052
L-iditol 2-dehydrogenase (EC 1.1.1.14) - Bacillus subtilis
N;Alternate names: glucitol dehydrogenase; polyol dehydrogenase; sorbitol dehydrogen
C;Species: Bacillus subtilis
C;Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 03-Nov-2000
C;Accession: A45052; E69638; I39862
R;Ng, K.; Ye, R.; Wu, X.C.; Wong, S.L.
J. Biol. Chem. 267, 24989-24994, 1992
A;Title: Sorbitol dehydrogenase from Bacillus subtilis. Purification, characterizati
A;Reference number: A45052; MUID:93094198; PMID:1460002
A;Accession: A45052
A;Molecule type: DNA
```

A;Residues: 1-353 <NG1>
A;Cross-references: GB:M96947; NID:g304152; PIDN:AAA22508.1; PID:g304153
A;Experimental source: strain W168
A;Note: sequence extracted from NCBI backbone (NCBIN:120108, NCBIP:120111)
A;Note: sequencing of the amino end of the mature protein indicated that Met-1 is remove
A;Note: presents arguments that the third ligand of the catalytic zinc is 156-Glu
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69638
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-353 <KUN>
A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12434.1; PID:g2632928
A;Experimental source: strain W168
R;Ye, R.; Wong, S.L.
J. Bacteriol. 176, 3314-3320, 1994
A;Title: Transcriptional regulation of the Bacillus subtilis glucitol dehydrogenase gene
A;Reference number: I39862; MUID:94253000; PMID:8195086
A;Accession: I39862
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <YER>
A;Cross-references: GB:L16626; NID:g436962; PIDN:AAA20875.1; PID:g530793
C;Genetics:
A;Gene: gutB
C;Complex: homotetramer
C;Function:
A;Description: catalyzes the oxidation by NAD+ of sorbitol to fructose
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: homotetramer; NAD; oxidoreductase; zinc
F;30-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;175-204/Region: beta-alpha-beta NAD nucleotide-binding fold
F;45,70,156/Binding site: zinc, catalytic (Cys, His, Glu) #status predicted
F;100,103,106,114/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 20.3%; Score 408.5; DB 1; Length 353;
Best Local Similarity 27.9%; Pred. No. 8.9e-24;
Matches 109; Conservative 66; Mismatches 155; Indels 61; Gaps 11;

```
Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          ||  : : | :|:  ||:| |:| | ::| |||   ||||:||  :              |
Db      9 MKAAVMHNTREIKI-ETLPVPDI-NHDEVLIKVMAVGICGSDLHYY------------T 53

Qy     61 KDKISGY--ELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQD 118
          :|  |  | |   ||| :| :  ||| |   | |||||||    |
Db     54 NGRIGNYVVEKPFILGHECAGEIAAVGSSVDQFKVGDRVAVEPGVTC------------ 100

Qy    119 LGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPIS 178
          | | ||: |  | |  : |       | | :|:   :| :  :|||:  :  ||:|| |
Db    101 -GRCEACKEGRYNLCPDVQFLATPPVDGAFVQYIKMRQDFVFLIPDSLSYEEAALIEPFS 159

Qy    179 VAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE 238
          |  ||  | : ||| |  ::| ||:||  : | :   || |: ::   :| : ||::||
Db    160 VGIHAAARTKLQPGSTIAIMGMGPVGLMAVAAAKAFGAGTIIVTDLEPLRLEAAKKMGAT 219

Qy    239 VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSG------------ 286
                ||   :|  :  : |   |:: :| |    ::: :    |
Db    220 HIINIREQDALEEIKT-ITNDRGVDVAWETAGNPAALQSALASVRRGGKLAIVGLPSQNE 278

Qy    287 IAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMIT 346
          | :||    |: |    :    |   | | |            :: |  |::     : ::|
Db    279 IPLNVPFIADNEIDIYGI-FRYANTYPKG------------IEFLASGIVD---TKHLVT 322

Qy    347 GKVHLKDGVEKGFKQLIEHKENNVKILVTPN 377
          :  |:   :   :: :: |   :|::| ||
Db    323 DQYSLEQ-TQDAMERALQFKNECLKVMVYPN 352
```

RESULT 13
T23889
hypothetical protein R04B5.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C;Accession: T23889
R;Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19814
A;Accession: T23889
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-347 <WIL>
A;Cross-references: EMBL:Z70782; PIDN:CAA94841.1; GSPDB:GN00023; CESP:R04B5.5
A;Experimental source: clone R04B5
C;Genetics:
A;Gene: CESP:R04B5.5
A;Map position: 5
A;Introns: 62/3; 138/2; 221/2; 259/3; 300/2
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 19.2%; Score 386.5; DB 2; Length 347;
Best Local Similarity 29.6%; Pred. No. 4.3e-22;
Matches 112; Conservative 62; Mismatches 151; Indels 53; Gaps 12;

```
Qy      7 YGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISG 66
            || :|:|  | || |:   || | :||   ||||:|:  :|:      |  |:
Db     11 YGVDDLRL-EQVPIPK-PGPNQVLVKVHTVGICGSDVHYWTHGAIGPFVVKE-------- 60

Qy     67 YELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQ 126
             |:  ||| || | |||:  |  :| |||:|:|  |                 ||  |:
Db     61 ---PMIVGHETSGIVSEVGNEVKHLKVGDRIAMEPGLPCK-------------LCEHCK 103

Qy    127 SGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVER 186
          :|  | |  : |       |  : :||:  |   ||||::  : |||:||:||| ||  |
Db    104 TGRYNLCPEMRFFATPPVHGTLSRFVVHDADFCFKLPDNLSFEDGALIEPLSVAIHACRR 163

Qy    187 ARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDPSTCD 246
            | |   |||| ||||:  ::  :   |||:| ::    |  ||:|||:     :| :
Db    164 GNVQMGHRVLVLGAGPIGVLNLITAKAVGAGKVVITDLDDGRLALAKKLGAD----ATIN 219

Qy    247 DANAVLKAMVPE------NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDH--- 297
              | |:  |       ::       :|:|   :  |:|  |   |: | | :  |
Db    220 VKGKSLDAVKSEIITALGDQQPDVCIECTGAQPSIETAITTTKSGGVIVLVGLGADRVEI 279

Qy    298 PIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLK-DGVE 356
          ||    :   :|   |  | |  :   :: :  | ::|         : | | :  :
Db    280 PI----IESATREVDMRGIFRY-VNCYPTAIELISSGKLNLSGLT-----RAHYKLEETQ 329

Qy    357 KGFKQLIEHKENNVKILV 374
          : ||:    | : :|: :
Db    330 EAFKR--TQKADVIKVFI 345
```

RESULT 14
S32484
L-iditol 2-dehydrogenase (EC 1.1.1.14) - silkworm
N;Alternate names: sorbitol dehydrogenase
C;Species: Bombyx mori (silkworm)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jun-2000
C;Accession: S32484
R;Niimi, T.; Yamashita, O.; Yaginuma, T.
Eur. J. Biochem. 213, 1125-1131, 1993
A;Title: A cold-inducible Bombyx gene encoding a protein similar to mammalian sorbito

```
A;Reference number: S32484; MUID:97087160; PMID:8504807
A;Accession: S32484
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-348 <NII>
A;Cross-references: EMBL:D13371; NID:g217259; PIDN:BAA02634.1; PID:g217260
C;Complex: homotetramer
C;Function:
A;Description: catalyzes the oxidation by NAD+ of sorbitol to fructose
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: homotetramer; metalloprotein; NAD; oxidoreductase; zinc
F;25-337/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;170-199/Region: beta-alpha-beta NAD nucleotide-binding fold
F;40,65,151/Binding site: zinc, catalytic (Cys, His, Gln) #status predicted
F;95,98,101,109/Binding site: zinc, noncatalytic (Cys) #status predicted

  Query Match             18.4%;  Score 370.5;  DB 1;  Length 348;
  Best Local Similarity   30.7%;  Pred. No. 7.2e-21;
  Matches  103;  Conservative   56;  Mismatches  146;  Indels   31;  Gaps   10;

Qy       7 YGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISG 66
Db      10 HGANDVRI-EKIPVPEI-NDDEVLIKIDCVGICGSDVK--LYSTGTC------GADVI-- 57

Qy      67 YELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQ 126
Db      58 -DKPIVIGHEGAGTVVKVGDKVSSLRVGDRVAIEPTQPC--RS------------CELCK 102

Qy     127 SGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVER 186
Db     103 RGKYNLCVEPRYCSSMGAPGNLCRYYKHVADFCHKLPDNLTMEEGAAVQPLAIVIHACNR 162

Qy     187 ARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAE--VFDPST 244
Db     163 AKITLGSKIVILGAGPIGILCAMSAKAMGASKIILTDVVQSRLDAALELGADNVLLVRRE 222

Qy     245 CDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPM 304
Db     223 YTDEEVVEKIVKLLGDRPDVSIDACGYGSAQRVALLVTKTAGLVLVVGI-ADKTVE-LPL 280

Qy     305 SLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDK 340
Db     281 SQALLREVDVVGSFRIMNTYQPALAAVSSGAIPLDK 316


RESULT 15
B72381
alcohol dehydrogenase, zinc-containing - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: B72381
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: B72381
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-395 <ARN>
A;Cross-references: GB:AE001720; GB:AE000512; NID:g4980906; PIDN:AAD35497.1; PID:g498091
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0412
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

  Query Match             18.4%;  Score 370;  DB 2;  Length 395;
  Best Local Similarity   34.0%;  Pred. No. 9.2e-21;
  Matches  109;  Conservative   37;  Mismatches  111;  Indels   64;  Gaps   13;

Qy       4 LLYYGTNDIRYSET----VPEPEIKNPNDVKIKVSYCGICGTDL-KEFTYSGGPVFFPKQ 58
Db      27 LTWLGSKVWRYPEVRVEEVPEPRIEKPTEIIIKVKACGICGSDVHMAQTDEEGYILYP-- 84

Qy      59 GTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKP------GDRVAVEATSHCSDRSRYK 112
Db      85 ----GLTGF--PVTLGHEFSGVVVEAGPEAINRRTNKRFEIGEPVCAEEMLWC------- 131

Qy     113 DTVAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYV--------------VYGED 157
Db     132 -------GHCRPCAEGFPNHCENLNELGF-NVDGAFAEYVKVDAKYAWSLRELEGVYEGD 183

Qy     158 HMVKLPDSIPDDIGALVEPISVAWHA--VERARFQPGQTALVLGGGPIGLATILALQGHH 215
Db     184 RLFL--------AGSLVEPTSVAYNAVIVRGGGIRPGDNVVILGGGPIGLAAVAILKHAG 235

Qy     216 AGKIVCSEPALIRRQFAKELGAE-VFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQT 274
Db     236 ASKVILSEPSEVRRNLAKELGADHVIDPTKENFVEAVLD--YTNGLGAKLFLEATGVPQL 293

Qy     275 FTTSI--VATGPSGIAVNVAV 293
Db     294 VWPQIEEVIWRARGINATVAI 314


Search completed: May  2, 2003, 13:01:43
Job time : 623 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 2, 2003, 12:08:47 ; Search time 239 Seconds
(without alignments)
46.781 Million cell updates/sec

Title: US-10-020-674-2
Perfect score: 2013
Sequence: 1 MKGLLYYGTNDIRYSETVPE.....QLIEHKENNVKILVTPNEVS 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	491.5	24.4	354	4	US-09-134-001C-3374	Sequence 3374, Ap
2	361	17.9	371	4	US-09-347-803-27	Sequence 27, Appl
3	360.5	17.9	356	3	US-08-466-548B-2	Sequence 2, Appli
4	360.5	17.9	356	3	US-07-998-226F-2	Sequence 2, Appli
5	360.5	17.9	356	5	PCT-US93-12560-2	Sequence 2, Appli
6	349	17.3	316	4	US-09-347-803-20	Sequence 20, Appl
7	327	16.2	301	4	US-09-347-803-22	Sequence 22, Appl
8	326	16.2	341	4	US-09-347-803-16	Sequence 16, Appl
9	266.5	13.2	382	4	US-09-134-001C-4483	Sequence 4483, Ap
10	265.5	13.2	381	4	US-09-134-001C-3003	Sequence 3003, Ap
11	244.5	12.1	336	1	US-08-713-254-2	Sequence 2, Appli
12	244.5	12.1	336	4	US-08-855-767-2	Sequence 2, Appli
13	236	11.7	352	4	US-09-504-358-18	Sequence 18, Appl
14	236	11.7	352	4	US-09-954-314-18	Sequence 18, Appl
15	233.5	11.6	374	4	US-09-347-878-52	Sequence 52, Appl
16	229.5	11.4	374	2	US-07-857-224B-82	Sequence 82, Appl
17	228.5	11.4	374	2	US-07-857-224B-84	Sequence 84, Appl
18	225.5	11.2	376	1	US-08-608-241-2	Sequence 2, Appli
19	225.5	11.2	376	2	US-08-922-182-2	Sequence 2, Appli
20	225.5	11.2	376	2	US-08-919-953-2	Sequence 2, Appli
21	225.5	11.2	376	4	US-09-192-983-2	Sequence 2, Appli
22	225.5	11.2	386	4	US-09-370-838-81	Sequence 81, Appl
23	223.5	11.1	374	2	US-07-857-224B-83	Sequence 83, Appl
24	222.5	11.1	375	4	US-09-347-878-56	Sequence 56, Appl
25	222	11.0	346	2	US-07-857-224B-96	Sequence 96, Appl
26	222	11.0	375	2	US-07-857-224B-86	Sequence 86, Appl
27	221.5	11.0	377	2	US-07-857-224B-90	Sequence 90, Appl
28	218	10.8	344	4	US-09-134-001C-5343	Sequence 5343, Ap
29	214	10.6	374	2	US-07-857-224B-85	Sequence 85, Appl
30	213.5	10.6	376	2	US-07-857-224B-91	Sequence 91, Appl
31	208.5	10.4	349	2	US-07-857-224B-95	Sequence 95, Appl
32	206	10.2	374	4	US-09-347-878-54	Sequence 54, Appl
33	204	10.1	379	2	US-07-857-224B-87	Sequence 87, Appl
34	201.5	10.0	374	2	US-07-857-224B-80	Sequence 80, Appl
35	201.5	10.0	374	2	US-07-857-224B-81	Sequence 81, Appl
36	195.5	9.7	376	2	US-07-857-224B-89	Sequence 89, Appl
37	194.5	9.7	352	2	US-08-805-191-2	Sequence 2, Appli
38	190.5	9.5	376	2	US-07-857-224B-88	Sequence 88, Appl
39	189.5	9.4	343	4	US-09-363-189B-4	Sequence 4, Appli
40	189	9.4	379	3	US-08-937-610-2	Sequence 2, Appli
41	185.5	9.2	344	2	US-07-857-224B-93	Sequence 93, Appl
42	179.5	8.9	345	2	US-07-857-224B-94	Sequence 94, Appl
43	170.5	8.5	344	2	US-07-857-224B-92	Sequence 92, Appl
44	148	7.4	93	4	US-08-936-165A-466	Sequence 466, App
45	147.5	7.3	396	4	US-08-860-656B-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-3374
; Sequence 3374, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3374
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3374

Query Match 24.4%; Score 491.5; DB 4; Length 354;
Best Local Similarity 34.9%; Pred. No. 3.5e-43;
Matches 132; Conservative 59; Mismatches 154; Indels 33; Gaps 11;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||  ::||  |:|  :   ||:    |:|:::|||: ||||||| |  |  ||:|
Db       5 MKAAVWYGQKDVRVEDR--EPKAIKDNEVQVKVSWAGICGTDLHE--YLEGPIFI-STDQ 59

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             | : |   |:  |||||| |  ||  |:  | |||| |  |'  | | : ::    |
Db      60 PDPLLGQTAPVTLGHEFSGVVENVGKNVSRFKKGDRVVVNPT--VSKREKPENVDLYD-- 115

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
                          || || |: | |||:    | ::  |||::    |||||| :||
Db     116 -------------GYSFIGL-GSDGAFAEFTNAPETNVYHLPDNVSAREGALVEPTAVA 160

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
             ||:      | |  | | ||||| ||:| :   | ||   : :  |   || :||  |
Db     161 VQAVKEGELLFGDTVAVFGAGPIGLLTIVAAKAAGASKIFVFDLSEERLAKAKSVGATHV 220

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           ::    |    | :     : |   :|: :||  |   ||  | | | || |:::| ||:
Db     221 YNSGNVDPVQTVYE---HTDNGVDVSFEVAGVGITLQQSIEVTRPRGTAVIVSIFG-HPV 276

Qy     300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
           | |:    :    | :: ||   ||: :  : :| :::   : ::| :: | : || ||
Db     277 EFNPLLQMNKGVKLTTTIAYTPTTFQQTIDLIANGSLNV---KDVVTDQIELDNIVESGF 333
```

```
Qy     360 KQLIEHKENNVKILVTPN 377
           ||:  | :  |||| |
Db     334 NQLVNDK-SQAKILVRLN 350


RESULT 2
US-09-347-803-27
; Sequence 27, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Hitz, Bill
;  APPLICANT: Kinney, Tony
;  APPLICANT: Orozco, Buddy
;  TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
;  FILE REFERENCE: BB-1176
;  CURRENT APPLICATION NUMBER: US/09/347,803
;  CURRENT FILING DATE: 1999-07-02
;  EARLIER APPLICATION NUMBER: 60/092,952
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 27
;   LENGTH: 371
;   TYPE: PRT
;   ORGANISM: Malus domestica
US-09-347-803-27

  Query Match             17.9%;  Score 361;  DB 4;  Length 371;
  Best Local Similarity   35.6%;  Pred. No. 2e-29;
  Matches   94;  Conservative   33;  Mismatches  109;  Indels   28;  Gaps    4;

Qy      26 PNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVVEVG 85
           |:||::::   ||||:|:  |       |  |:          |:  ||| :| : |||
Db      51 PHDVRVRLRAVGICGSDVHHFKNMRCVDFIVKE----------PMVIGHECAGIIEEVG 99

Qy      86 SGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLSFCGLGGAS 145
           | |  : ||||||:|    |              |  |: |  | |  : | |    :
Db     100 SEVEHLVPGDRVALEPGISCK-------------RCNLCKQGRYNLCRKMKFFGSPPNN 145

Qy     146 GGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTALVLGGGPIGL 205
           |  |  ||:  |   ||||::  : ||: ||:||  ||  ||        ||:| |||||
Db     146 GCLANQVVHPGDLCFKLPDNVSLEEGAMCEPLSVGIHACRRANVCQETNVLVVGAGPIGL 205

Qy     206 ATILALQGHHAGKIVCSEPALIRRQFAKELGAE-VFDPSTC--DDANAVLKAMVPENEGF 262
           |:|| :   | :|| ::    |   || |||: |   ||   | |  | |       |
Db     206 VTLLAARAFGAPRIVIADVNDERLLIAKSLGADAVVKVSTNIEDVAEEVAKIQKVLENGV 265

Qy     263 HAAFDCSGVPQTFTTSIVATGPSG 286
           |||:|  :| ||:: || | |
Db     266 DVTFDCAGFNKTITTALSATRPGG 289


RESULT 3
US-08-466-548B-2
; Sequence 2, Application US/08466548B
; Patent No. 6022712
;  GENERAL INFORMATION:
;    APPLICANT:  sarthy, aparna v
;    APPLICANT:  schopp, cynthia w
;    TITLE OF INVENTION:  ENHANCED YEAST EXPRESSION USING
;    TITLE OF INVENTION:  REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
;    TITLE OF INVENTION:  DEHYDROGENASE GENE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LAORTORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,548B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/998,226
;      FILING DATE:  30-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WALSH, ANDREA C
;      REGISTRATION NUMBER:  34988
;      REFERENCE/DOCKET NUMBER:  5283.us.01
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  356 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-466-548B-2

  Query Match             17.9%;  Score 360.5;  DB 3;  Length 356;
  Best Local Similarity   30.3%;  Pred. No. 2.1e-29;
  Matches  111;  Conservative   56;  Mismatches  160;  Indels   39;  Gaps   11;

Qy      16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYEL--PLCP 73
           |  | | ||:|: ||: :   ||||:|:  :  |||            |  | |  |:
Db      20 EQRPIPTIKDPHYVKLAIKATGICGSDI-HYYRSGG-----------IGKYILKAPMVL 66

Qy      74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
           ||| || |||||  || || |||||:|        ||| |           : |  | |
Db      67 GHESSGQVVEVGDAVTRVKVGDRVAIEP----GVPSRYSDET----------KEGRYNLC 112

Qy     134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
             ::|       |   :| :  || :||||: :  : || |||:||  |: : |  : |
Db     113 PHMAFAATPPIDGTLVKYYLSPEDFLVKLPEGVSYEEGACVEPLSVGVHSNKLAGVRFGT 172

Qy     194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPS--TCDDANA 250
             :| | ||:|| |    :  |  ::  :   : | ||: ||   |:.|  : | |
Db     173 KVVVFGAGPVGLLTGAVARAFGATDVIFVDVFDNKLQRAKDFGATNTFNSSQFSTDKAQD 232

Qy     251 VLKAMVPENEGFHA--AFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTY 308
           :   :     | ||   |:|||       ::  |   |  | | : | :   |    ::
Db     233 LADGVQKLLGGNHADVVFECSGADVCIDAAVKTTKVGGTMVQVGM-GKNYTNFPIAEVSG 291

Qy     309 QEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKEN 368
           :|    |   |: |::: | :  | :::   : :|| |   :|   | :   | |
Db     292 KEMKLIGCFRYSFGDYRDAVNLVATGKVNV---KPLITHKFKFEDAA-KAYDYNIAHGGE 347

Qy     369 NVKILV 374
             || ::
Db     348 VVKTII 353


RESULT 4
US-07-998-226F-2
; Sequence 2, Application US/07998226F
; Patent No. 6033898
;  GENERAL INFORMATION:
;    APPLICANT:  sarthy, aparna v
;    APPLICANT:  schopp, cynthia w
;    TITLE OF INVENTION:  ENHANCED YEAST EXPRESSION USING
;    TITLE OF INVENTION:  REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
;    TITLE OF INVENTION:  DEHYDROGENASE GENE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LAORTORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  US
;      ZIP:  60064-3500
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/998,226F
;      FILING DATE:  30-DEC-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WALSH, ANDREA C
;      REGISTRATION NUMBER:  34988
;      REFERENCE/DOCKET NUMBER:  5283.us.01
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  356 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-998-226F-2

  Query Match             17.9%;  Score 360.5;  DB 3;  Length 356;
  Best Local Similarity   30.3%;  Pred. No. 2.1e-29;
  Matches  111;  Conservative   56;  Mismatches  160;  Indels   39;  Gaps   11;

Qy      16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYEL--PLCP 73
           |  | | ||:|: ||: :   ||||:|:  :  |||           |  | |  |:
Db      20 EQRPIPTIKDPHYVKLAIKATGICGSDI-HYYRSGG-----------IGKYILKAPMVL 66

Qy      74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
           ||| || |||||  || || |||||:|        ||| |                : |  | |
Db      67 GHESSGQVVEVGDAVTRVKVGDRVAIEP----GVPSRYSDET----------KEGRYNLC 112

Qy     134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
            ::|       |   :| :  || :||||: :  : || |||:||  |: : |  : |
Db     113 PHMAFAATPPIDGTLVKYYLSPEDFLVKLPEGVSYEEGACVEPLSVGVHSNKLAGVRFGT 172

Qy     194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPS--TCDDANA 250
             :| | ||:|| |    :   |  ::  :    : | ||: ||   |: |. : | |
Db     173 KVVVFGAGPVGLLTGAVARAFGATDVIFVDVFDNKLQRAKDFGATNTFNSSQFSTDKAQD 232

Qy     251 VLKAMVPENEGFHA--AFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTY 308
           :   :     | ||   |:|||       ::  |   |  | | : |   |    ::
Db     233 LADGVQKLLGGNHADVVFECSGADVCIDAAVKTTKVGGTMVQVGM-GKNYTNFPIAEVSG 291

Qy     309 QEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKEN 368
           :|    |   |:  |::: |  :  | :::   : :|| |   :|   | :   | |
Db     292 KEMKLIGCFRYSFGDYRDAVNLVATGKVNV---KPLITHKFKFEDAA-KAYDYNIAHGGE 347

Qy     369 NVKILV 374
             || ::
Db     348 VVKTII 353


RESULT 5
PCT-US93-12560-2
; Sequence 2, Application PC/TUS9312560
;  GENERAL INFORMATION:
;    APPLICANT:  sarthy, aparna v
;    APPLICANT:  schopp, cynthia w
;    TITLE OF INVENTION:  ENHANCED YEAST EXPRESSION IN USING
;    TITLE OF INVENTION:  REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
;    TITLE OF INVENTION:  DEHYDROGENASE GENE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LAORTORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/12560
;      FILING DATE:  22-DEC-1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/998,226
;      FILING DATE:  30-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WALSH, ANDREA C
;      REGISTRATION NUMBER:  34988
;      REFERENCE/DOCKET NUMBER:  5283.us.01
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  356 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-12560-2

  Query Match             17.9%;  Score 360.5;  DB 5;  Length 356;
  Best Local Similarity   30.3%;  Pred. No. 2.1e-29;
  Matches  111;  Conservative   56;  Mismatches  160;  Indels   39;  Gaps   11;

Qy      16 ETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYEL--PLCP 73
           |  | | ||:|: ||: :   ||||:|:  :  |||           |  | |  |:
Db      20 EQRPIPTIKDPHYVKLAIKATGICGSDI-HYYRSGG-----------IGKYILKAPMVL 66

Qy      74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
           ||| || |||||  || || |||||:|        ||| |                : |  | |
Db      67 GHESSGQVVEVGDAVTRVKVGDRVAIEP----GVPSRYSDET----------KEGRYNLC 112

Qy     134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
            ::|       |   :| :  || :||||: :  : || |||:||  |: : |  : |
Db     113 PHMAFAATPPIDGTLVKYYLSPEDFLVKLPEGVSYEEGACVEPLSVGVHSNKLAGVRFGT 172

Qy     194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPS--TCDDANA 250
             :| | ||:|| |    :   |  ::  :    : | ||: ||   |: |  : | |
Db     173 KVVVFGAGPVGLLTGAVARAFGATDVIFVDVFDNKLQRAKDFGATNTFNSSQFSTDKAQD 232

Qy     251 VLKAMVPENEGFHA--AFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTY 308
           :   :     | ||   |:|||       ::  |   |  | | : |   |    ::
Db     233 LADGVQKLLGGNHADVVFECSGADVCIDAAVKTTKVGGTMVQVGM-GKNYTNFPIAEVSG 291

Qy     309 QEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKEN 368
           :|    |   |:  |::: |  :  | :::   : :|| |   :|   | :   | |
Db     292 KEMKLIGCFRYSFGDYRDAVNLVATGKVNV---KPLITHKFKFEDAA-KAYDYNIAHGGE 347

Qy     369 NVKILV 374
             || ::
Db     348 VVKTII 353


RESULT 6
US-09-347-803-20
; Sequence 20, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Hitz, Bill
;  APPLICANT: Kinney, Tony
;  APPLICANT: Orozco, Buddy
;  TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
;  FILE REFERENCE: BB-1176
;  CURRENT APPLICATION NUMBER: US/09/347,803
;  CURRENT FILING DATE: 1999-07-02
;  EARLIER APPLICATION NUMBER: 60/092,952
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
```

```
;   LENGTH: 316
;   TYPE: PRT
;   ORGANISM: Glycine max
US-09-347-803-20

  Query Match             17.3%;  Score 349;  DB 4;  Length 316;
  Best Local Similarity   33.3%;  Pred. No. 2.8e-28;
  Matches   88;  Conservative   36;  Mismatches  112;  Indels   28;  Gaps    4;

Qy      26 PNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVVEVG 85
           |:||::::   ||||:|:          |  |:           |:  ||| :| : |||
Db      41 PHDVRVRMKAVGICGSDVHYLKTLRCAHFIVKE----------PMVIGHECAGIIEEVG 89

Qy      86 SGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLSFCGLGGAS 145
           | | |: ||||||:|    |                |  |: |  | |  : |
Db      90 SQVKSLVPGDRVAIEPGISCWH-------------CNHCKHGRYNLCDDMKFFATPPVH 135

Qy     146 GGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTALVLGGGPIGL 205
           |  |  :|:  |   ||||::  : ||: ||:||  ||  ||    |    |::| |||||
Db     136 GSLANQIVHPADLCFKLPDNVSLEEGAMCEPLSVGVHACRRANIGPETNVLIMGAGPIGL 195

Qy     206 ATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPST--CDDANAVLKAMVPENEGF 262
            |:|| :   | | |  :    |   || ||| ::   ||   | |  |::       |
Db     196 VTMLAARAFGAPKTVIVDVDDHRLSVAKSLGADDIIKVSTNIKDVAEEVVQIQKVMGAGI 255

Qy     263 HAAFDCSGVPQTFTTSIVATGPSG 286
           |||:|  :| : :: || | |
Db     256 DVTFDCAGFDKTMSPALSATQPGG 279


RESULT 7
US-09-347-803-22
; Sequence 22, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Hitz, Bill
;  APPLICANT: Kinney, Tony
;  APPLICANT: Orozco, Buddy
;  TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
;  FILE REFERENCE: BB-1176
;  CURRENT APPLICATION NUMBER: US/09/347,803
;  CURRENT FILING DATE: 1999-07-02
;  EARLIER APPLICATION NUMBER: 60/092,952
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 301
;   TYPE: PRT
;   ORGANISM: Triticum aestivum
US-09-347-803-22

  Query Match             16.2%;  Score 327;  DB 4;  Length 301;
  Best Local Similarity   31.5%;  Pred. No. 5.4e-26;
  Matches   84;  Conservative   42;  Mismatches  107;  Indels   34;  Gaps    5;

Qy      26 PNDVKIKVSYCGICGTD---LKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVV 82
           | ||:::   ||||:|   |||   :    |  |:           |:  ||| :| :
Db      25 PYDVRVRMKAVGICGSDVHYLKEMRIAH---FVVKE----------PMVIGHECAGIIE 70

Qy      83 EVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLSFCGLG 142
           |||  ||   :  |||||:|    |                |  |: |  | |  : |
Db      71 EVGDGVKHLAVGDRVALEPGISC-------------WRCRHCKGGRYNLCDDMKFFATP 116

Qy     143 GASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTALVLGGGP 202
             |  |: :|:  |   ||||::  : ||: ||:||  ||  ||     :: |::| ||
Db     117 PYHGSLADQIVHPGDLCFKLPDNVSLEEGAMCEPLSVGVHACRRADVGAEKSVLIMGAGP 176

Qy     203 IGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEV---FDPSTCDDANAVLKAMVPEN 259
           ||| |:|: :   | :|| ::    |   || |||:       :| | |  : :
Db     177 IGLVTMLSARAFGAPRIVIADVDDHRLSVAKSLGADAVVKVSGNTEDLAGEIERIQAAMG 236

Qy     260 EGFHAAFDCSGVPQTFTTSIVATGPSG 286
             : ||:|  :| :|:: || | |
Db     237 GDIDVSLDCAGFSKTMSTALEATRPGG 263


RESULT 8
US-09-347-803-16
; Sequence 16, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Hitz, Bill
;  APPLICANT: Kinney, Tony
;  APPLICANT: Orozco, Buddy
;  TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
;  FILE REFERENCE: BB-1176
;  CURRENT APPLICATION NUMBER: US/09/347,803
;  CURRENT FILING DATE: 1999-07-02
;  EARLIER APPLICATION NUMBER: 60/092,952
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-347-803-16

  Query Match             16.2%;  Score 326;  DB 4;  Length 341;
  Best Local Similarity   29.7%;  Pred. No. 8.4e-26;
  Matches  103;  Conservative   47;  Mismatches  137;  Indels   60;  Gaps   10;

Qy      26 PNDVKIKVSYCGICGTD---LKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVV 82
           | ||:::   ||||:|   |:|   :    |  |:           |:  ||| :| |
Db      33 PYDVRVRMKAVGICGSDVHYLREMRIAH---FVVKE----------PMVIGHECAGVVE 78

Qy      83 EVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLSFCGLG 142
           |||:||| :  |||||:|    |                |  |: |  |   :: |
Db      79 EVGAGVTHLSVGDRVALEPGVSC-------------WRCRHCKGGRYNPVRNMKFFATP 124

Qy     143 GASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTALVLGGGP 202
             |  |  ||:  |   |||| :  : ||: ||:||  ||  ||    |    ||:| ||
Db     125 PVHGSLANQVVHPADLCFKLPDGVSLEEGAMCEPLSVGVHACRRAGVGPETGVLVVGAGP 184

Qy     203 IGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEV---FDPSTCDDANAVLKAMVPEN 259
           ||| ::|| :   | ::|  :    |   |: |||:       |   | |: | :
Db     185 IGLVSLLAARAFGAPRVVVVDVDDHRLAVARSLGADAAVRVSPRAEDLADEVERIRAAMG 244

Qy     260 EGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCY 319
             : ||:|  :| :|:: || | |    | :  :    | : |  :|    |   |
Db     245 SDIDVSLDCAGFSKTMSTALEATRPGGKVCLVGMGHNE----MTLPLAAREVDVVGVFRY 300

Qy     320 TVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHK 366
             ||            : :|  |   :||| :        | || |:
Db     301 --KDTWP---------LCIDFLR---SGKVDV--------KPLITHR 325


RESULT 9
US-09-134-001C-4483
; Sequence 4483, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
```

```
; SEQ ID NO 4483
;   LENGTH: 382
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4483

  Query Match             13.2%;  Score 266.5;  DB 4;  Length 382;
  Best Local Similarity   25.5%;  Pred. No. 1.8e-19;
  Matches   97;  Conservative   63;  Mismatches  145;  Indels   75;  Gaps   16;

Qy      20 EPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSG 79
           :| | :  |  |::    ||||||           |  | :  |   |    |||  |
Db      51 KPTITSSTDAIIRIVKTTICGTDL----------HIIKGDTPEVKSHTTL----GHEGIG 96

Qy      80 TVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASLSFC 139
            : |:|  | : | ||:| :   | |              | |  |: |        : |
Db      97 IIEEIGDNVNNFKVGDKVIISCISSC--------------GKCYYCKKG------IYAHC 136

Qy     140 GLGG-------ASGGFAEY--VVYGEDHMVKLPDSIPDDIGALVEPISVAWH--AVERAR 188
             ||        :|  |||  | : :: :   | :: ||   ::  |    :   | : :
Db     137 ENGGGWILGHLVNGTQAEYVKVPFADNSLYHAPSNLKDDALVMLSDILPTGYEIGVLKGK 196

Qy     189 FQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EVFDPSTCDD 247
           :|| |  ::| ||:||| :|  | :   ||:  :    | : ||||||  : :    :
Db     197 VKPGCTVAIVGAGPVGLAALLTAQFYSPSKIIMIDLDDNRLETAKELGATHLINSKETET 256

Qy     248 ANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLT 307
           |   :|:: |   |   | :  |:||||       | |   || | |      :|: |
Db     257 AIKKVKSLNP--RGVDVAIEAVGIPQTFDLCQNLIGVDGTIANVGVHG------LPVQLD 308

Qy     308 YQEKY------ATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG--- 358
           : :       || :  :    :|:::||:  :|   :  :::|    | : :|
Db     309 IDKLWIKNINVTTGLV--SGNTTEELLEALKSKII---QPEQLVTHYSKLSE-IESAYDL 362

Qy     359 FKQLIEHK------ENNVKI 372
           |:   :||      ||:: |
Db     363 FRNATDHKAIKLIIENDITI 382


RESULT 10
US-09-134-001C-3003
; Sequence 3003, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3003
;   LENGTH: 381
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3003

  Query Match             13.2%;  Score 265.5;  DB 4;  Length 381;
  Best Local Similarity   24.4%;  Pred. No. 2.3e-19;
  Matches   96;  Conservative   54;  Mismatches  152;  Indels   91;  Gaps   13;

Qy      14 YSETVPEP----EIKNP--NDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGY 67
           | |: |      |: ||  ::| ||:   |:| :||                | : :
Db      23 YKESKPLKIEYLELDNPSEHEVLIKIHAAGLCHSDLSVI----------NGNRPR---- 67

Qy      68 ELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQS 127
           ||:  ||| || |::||  || |  || |       |              | |: |:
Db      68 PLPMALGHEASGEVIKVGKAVTRVSEGDHVVCTFIPSC-------------GKCIPCKE 113

Qy     128 GSPNCCASLSFCGLGG-------------------ASGGFAEYVVYGEDHMVKLPDSIP 167
           | |  | : :     |                    ||||| |  |: :||:   ||
Db     114 GRPALCENGAISNEKGEMLEGGMRLSNDEGKVYHHLGISGFAEYSVVSENSIVKIDKKIP 173

Qy     168 DD-----------IGALVEPISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHH 215
           :            |||:          |  |: : |    |:| | |||  |: :
Db     174 FERAAAFGCAIITGIGAV----------VNTAQIRSGSNVAVVGLGGIGLNAIIGAKLAG 223

Qy     216 AGKIVCSEPALIRRQFAKELGAEVFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTF 275
           | :|:  :    : : ||: ||     |:  | :  :|  :|   |   ||: :||
Db     224 ANEIIALDINEDKFELAKQFGATATFNSSDKDIDEQIKEYIP--GGVEYAFETAGVVPAM 281

Qy     276 TTSIVATGPSGIAVNVAVWGDHP---IGFMPMSLTYQEKYATGS---MCYTVKDFQEVVK 329
           :   |   | |   :   :|     |  ::|  :|:   ||    |   :|    |
Db     282 KVAYQITKRGGTTVTTGL--PNPKDNFSFPQVTLAAEERTIKGSYVGSCVPDRDIPRFVN 339

Qy     330 ALEDGLISLDKARKMITGKVHLKDGVEKGFKQL 362
             | :::|     :  :|   | : :|| :|
Db     340 LYNQGRLNIDS----LISEVITLDEINEGFDRL 368


RESULT 11
US-08-713-254-2
; Sequence 2, Application US/08713254
; Patent No. 5763236
;  GENERAL INFORMATION:
;    APPLICANT:  KOJIMA, TOMOKO
;    APPLICANT:  YAMAMOTO, HIROAKI
;    APPLICANT:  KAWADA, NAOKI
;    APPLICANT:  MATSUYAMA, AKINOBU
;    TITLE OF INVENTION:  NOVEL ENZYME, A METHOD TO PREPARE SAID
;    TITLE OF INVENTION:  ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
;    TITLE OF INVENTION:  CONTAINING SAID DNA SEGMENT AND A METHOD OF PREPARING
;    TITLE OF INVENTION:  OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/713,254
;      FILING DATE:  12-SEP-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/311,328
;      FILING DATE:  23-SEP-1994
;      APPLICATION NUMBER:  JP 5-261649
;      FILING DATE:  24-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-337191
;      FILING DATE:  28-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-181308
;      FILING DATE:  02-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5763236man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  6423-001-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  336 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-713-254-2

  Query Match             12.1%;  Score 244.5;  DB 1;  Length 336;
  Best Local Similarity   25.8%;  Pred. No. 3e-17;
  Matches   92;  Conservative   53;  Mismatches  146;  Indels   65;  Gaps   16;

Qy      31 IKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVVEVGSGVTS 90
           :||   |:| :||    | |          |    | :    ||| :||| ||  | :
Db      36 LKVDAVGLCHSDL-HVIYEG----------LDCGDNYVM----GHEIAGTVAAVGDDVIN 80

Qy      91 VKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASL--SFCGLGGASGGF 148
           | |||||    : |              | |  |:    | | :   : || |  ||:
Db      81 YKVGDRVACVGPNGC-------------GGCKYCRGAIDNVCKNAFGDWFGL-GYDGGY 125

Qy     149 AEY-VVYGEDHMVKLPDSIPDDIGAL-VEPISVAWHAVERARFQPGQTALVLGGGPIGLA 206
           :| :|    :: ::||::  |: |   : :   :||:: |:  |    |::| | :|
Db     126 QQYLLVTRPRNLSRIPDNVSADVAAASTDAVLTPYHAIKMAQVSPTSNILLIGAGGLGGN 185

Qy     207 TILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDPSTCDDANAVLKAMVPEN---EGFH 263
           |   :  | |:  :     |  ||:|||:           || : : ||:     |
Db     186 AIQVAKAFGA-KVTVLDKKKEARDQAKKLGAD----------AVYETL-PESISPGSFS 232

Qy     264 AAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKD 323
           | ||   |  ||        | |: : | : |   : |   |  :|    ||   |  |
Db     233 ACFDFVSVQATFDVCQKYVEPKGVIMPVGL-GAPNLSFNLGDLALREIRILGSFWGTTND 291

Qy     324 FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKENNV---KILVTP 376
           :|:| : :| :      | :     ||:       : ||   ||    ::: |
Db     292 LDDVLKLVSEGKV------KPVVRSAKLKE-----LPEYIEKLRNNAYEGRVVFNP 336


RESULT 12
US-08-855-767-2
; Sequence 2, Application US/08855767
; Patent No. 6255092
;  GENERAL INFORMATION:
;    APPLICANT:  KOJIMA, TOMOKO
;    APPLICANT:  YAMAMOTO, HIROAKI
;    APPLICANT:  KAWADA, NAOKI
;    APPLICANT:  MATSUYAMA, AKINOBU
;    TITLE OF INVENTION:  NOVEL ENZYME, A METHOD TO PREPARE SAID
;    TITLE OF INVENTION:  ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
;    TITLE OF INVENTION:  CONTAINING SAID DNA SEGMENT AND A METHOD OF PREPARING
;    TITLE OF INVENTION:  OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,767
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/311,328
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-337191
;      FILING DATE:  28-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-181308
;      FILING DATE:  02-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 6255092man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  6423-001-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  336 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-855-767-2

  Query Match             12.1%;  Score 244.5;  DB 4;  Length 336;
  Best Local Similarity   25.8%;  Pred. No. 3e-17;
  Matches   92;  Conservative   53;  Mismatches  146;  Indels   65;  Gaps   16;

Qy      31 IKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEFSGTVVEVGSGVTS 90
           :||   |:| :||    | |          |    | :    ||| :||| ||  | :
Db      36 LKVDAVGLCHSDL-HVIYEG----------LDCGDNYVM----GHEIAGTVAAVGDDVIN 80

Qy      91 VKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCASL--SFCGLGGASGGF 148
           | |||||    : |              | |  |:    | | :  : || |  ||:
Db      81 YKVGDRVACVGPNGC-------------GGCKYCRGAIDNVCKNAFGDWFGL-GYDGGY 125

Qy     149 AEY-VVYGEDHMVKLPDSIPDDIGAL-VEPISVAWHAVERARFQPGQTALVLGGGPIGLA 206
           :| :|    :: ::||::  |: |   : :   :||:: |:  |    |::| | :|
Db     126 QQYLLVTRPRNLSRIPDNVSADVAAASTDAVLTPYHAIKMAQVSPTSNILLIGAGGLGGN 185

Qy     207 TILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDPSTCDDANAVLKAMVPEN---EGFH 263
           |   :  | |:  :     |  ||:|||:           || : : ||:     |
Db     186 AIQVAKAFGA-KVTVLDKKKEARDQAKKLGAD----------AVYETL-PESISPGSFS 232

Qy     264 AAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKD 323
           | ||   |  ||        | |: : | : |   : |   |  :|    ||   |  |
Db     233 ACFDFVSVQATFDVCQKYVEPKGVIMPVGL-GAPNLSFNLGDLALREIRILGSFWGTTND 291

Qy     324 FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFKQLIEHKENNV---KILVTP 376
           :|:| : :| :      | :     ||:       : ||   ||    ::: |
Db     292 LDDVLKLVSEGKV------KPVVRSAKLKE-----LPEYIEKLRNNAYEGRVVFNP 336


RESULT 13
US-09-504-358-18
; Sequence 18, Application US/09504358
; Patent No. 6365376
; GENERAL INFORMATION:
;  APPLICANT: Rouviere, Pierre E.
;  APPLICANT: Brzostowicz, Patricia C.
;  TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA
;  FILE REFERENCE: BC1001 US NA
;  CURRENT APPLICATION NUMBER: US/09/504,358
;  CURRENT FILING DATE: 2000-02-15
;  EARLIER APPLICATION NUMBER: 60/120,702
;  EARLIER FILING DATE: 1999-February-19
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: Brevibacterium sp HCU
US-09-504-358-18

  Query Match             11.7%;  Score 236;  DB 4;  Length 352;
  Best Local Similarity   28.1%;  Pred. No. 2.6e-16;
  Matches  104;  Conservative   40;  Mismatches  154;  Indels   72;  Gaps   16;

Qy      20 EPEIKNP----NDVKIKVSYCGICGTD--LKEFTYSGGPVFFPKQGTKD-KISGYELPLC 72
```

```
Db     15 EIELDRPKPMGREVLLKVTHAGVCHTDTHVQDGGYDLG-----SRGTLDMSTRGVTYPCV 69

Qy     73 PGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNC 132

Db     70 MGHETVGEVVEVGEDVTDVAVGDTCLAFPWIGC-------------GECGKCAHGHENA 115

Qy    133 CASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGAL-VEPISVAWHAV------- 184

Db    116 CDNGRALGI-IQFGGFAEYLL-----------LPDQRYAIDVAGVDPAWAATLACSGVT 162

Qy    185 -------ERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA 237

Db    163 SYSSARKATATVNPDEPIGVMGVGGVGMMTVAALVALGHKNIIAIDVSDENLASAQELGA 222

Qy    238 EVFDPSTCDDANAVLKAMVPENEG-FHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGD 296

Db    223 TL----TVNSKNATSHDLVEAAGGQFIAIIDLVNTGDTVALAFDALSRAGKIVQVGLFGG 278

Qy    297 H---PIGFMPM-SLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLK 352

Db    279 EFVVPTAIMALKGLTLQGNYVG-----TVEEVREVVELARQG--SLPKL--PITGGTLNV 329

Qy    353 DGVEKGFKQL 362

Db    330 DGVNDGLERL 339


RESULT 14
US-09-954-314-18
; Sequence 18, Application US/09954314
; Patent No. 6465224
; GENERAL INFORMATION:
;  APPLICANT: Rouviere, Pierre E.
;  APPLICANT:  Brzostowicz, Patricia C.
;  TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES
;  FILE REFERENCE: BC1001 US NA
;  CURRENT APPLICATION NUMBER: US/09/954,314
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/120,702
;  PRIOR FILING DATE: 1999-February-19
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: Brevibacterium sp HCU
US-09-954-314-18

  Query Match           11.7%;  Score 236;  DB 4;  Length 352;
  Best Local Similarity  28.1%;  Pred. No. 2.6e-16;
  Matches  104;  Conservative   40;  Mismatches  154;  Indels   72;  Gaps   16;

Qy     20 EPEIKNP----NDVKIKVSYCGICGTD--LKEFTYSGGPVFFPKQGTKD-KISGYELPLC 72

Db     15 EIELDRPKPMGREVLLKVTHAGVCHTDTHVQDGGYDLG-----SRGTLDMSTRGVTYPCV 69

Qy     73 PGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNC 132

Db     70 MGHETVGEVVEVGEDVTDVAVGDTCLAFPWIGC-------------GECGKCAHGHENA 115

Qy    133 CASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGAL-VEPISVAWHAV------- 184

Db    116 CDNGRALGI-IQFGGFAEYLL-----------LPDQRYAIDVAGVDPAWAATLACSGVT 162

Qy    185 -------ERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA 237

Db    163 SYSSARKATATVNPDEPIGVMGVGGVGMMTVAALVALGHKNIIAIDVSDENLASAQELGA 222

Qy    238 EVFDPSTCDDANAVLKAMVPENEG-FHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGD 296

Db    223 TL----TVNSKNATSHDLVEAAGGQFIAIIDLVNTGDTVALAFDALSRAGKIVQVGLFGG 278

Qy    297 H---PIGFMPM-SLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLK 352

Db    279 EFVVPTAIMALKGLTLQGNYVG-----TVEEVREVVELARQG--SLPKL--PITGGTLNV 329

Qy    353 DGVEKGFKQL 362

Db    330 DGVNDGLERL 339


RESULT 15
US-09-347-878-52
; Sequence 52, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
;  APPLICANT: Yuan, Chong
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
;  FILE REFERENCE: 25885-1651
;  CURRENT APPLICATION NUMBER: US/09/347,878C
;  CURRENT FILING DATE: 1999-07-06
;  NUMBER OF SEQ ID NOS: 75
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-347-878-52

  Query Match           11.6%;  Score 233.5;  DB 4;  Length 374;
  Best Local Similarity  25.3%;  Pred. No. 5.2e-16;
  Matches   96;  Conservative   57;  Mismatches  142;  Indels   85;  Gaps   17;

Qy     17 TVPEPEIKNP--NDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPG 74

Db     23 SIEEIEVAPPKTKEVRIKILATGICRTD--DHVIKGTMV-------------SKFPVIVG 67

Qy     75 HEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCCA 134

Db     68 HEATGIVESIGEGVTTVKPGDKVIPLFLPQCRE--------------CNACRNPDGNLCI 113

Qy    135 SLSFCGLGGASGG------------------FAEYVVYGEDHMVKLPDSIPDDIGALVE 175

Db    114 RSDITGRGVLADGTTRFTCKGKPVHHFMNTSTFTEYTVVDESSVAKIDDAAPPEKVCLIG 173

Qy    176 -PISVAW-HAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAK 233

Db    174 CGFSTGYGAAVKTGKVKPGSTCVVFGLGGVGLSVIMGCKSAGASRIIGID--LNKDKFEK 231

Qy    234 EL---GAEVFDP-STCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPS-GIA 288

Db    232 AMAVGATECISPKDSTKPISEVLSEMTGNNVGY--TFEVIGHLETMIDALASCHMNYGTS 289

Qy    289 VNVAVWGDHPIGFMPMS---LTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMI 345

Db    290 VVVGV---------PPSAKMLTYDP-----MLLFTGRTWKGCVFG---GLKSRDDVPKLV 332

Qy    346 TGKVHLKDGVEKGFKQLIEH 365

Db    333 TEFLAKKFDLD----QLITH 348


Search completed: May  2, 2003, 13:05:55
Job time : 240 secs
```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11975
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11975

Query Match 22.6%; Score 455.5; DB 10; Length 352;
Best Local Similarity 33.5%; Pred. No. 3.9e-34;
Matches 127; Conservative 55; Mismatches 162; Indels 35; Gaps 12;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || | ::   |:| ||   | :   | :|:::|:|||||||:|| |  |  ||   | |
Db       1 MKALRWHAARDLRLSEL--ERQAPRPGEVELEVAYCGICGSDLHE--YQSGPHSIP-QAE 55

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             :||   ||  |||| | |  :| ||   : |||||||    |              |
Db      56 AHPLSGCRAPLTLGHEFCGVVAALGPGVEGPRIGDRVAVEPEYRC----------G 101

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           |  |: |  | | |: | || |  |||||        : :||::     |::|| :||
Db     102 ECRYCREGRYNLCESMGFIGLMG-DGGFAERARVPAYMLHRLPDAVGFRQAAVLEPAAVA 160

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
           ||: |:   |||   | | |||||  ::  :      |  : :  |  | | ||
Db     161 LHALRRSSLAPGQRCAVFGLGPIGLLLVMLARLRGIEDIAAVDVSPERLALAGEFGASRA 220

Qy     240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
           |    | |     | : |      ||: :|   :   :: :    |  | |:: |:  :
Db     221 LDARDGDTA-----ARLREGGALDCAFEAAGSQASLDAALASLRKGGELVLVSLMGE--V 273

Qy     300 GFMPMSLTYQEKYATGSMCYTVKD-FQEVVKALEDGLISLDKARKMITGKVHLKDGVEKG 358
              |  :|    ||: |  :| : |::  | || : | :|   :|  | |:  || |
Db     274 RLDAFDLVNRELRLLGSVGY--RDAYPELIALLADGRLDLARA---VTRSVPLEQAVEHG 328

Qy     359 FKQLIEHKENNVKILVTPN 377
           |: |:  | : :|:|| ||
Db     329 FEALLRDK-SQLKVLVNPN 346
```

RESULT 5
US-09-908-744-53
; Sequence 53, Application US/09908744
; Publication No. US20030068791A1

; GENERAL INFORMATION:
; APPLICANT: Miasnikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelainen Mira
; APPLICANT: Gros H+kan
; APPLICANT: Toivari Mervi
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koivuranta Kari
; APPLICANT: Londesborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Menut Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/FI01/00051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03
; PRIOR APPLICATION NUMBER: US 08/110,672
; PRIOR FILING DATE: 1993-08-24
; PRIOR APPLICATION NUMBER: US 07/973,325
; PRIOR FILING DATE: 1992-11-05
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 53
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-908-744-53

Query Match 17.7%; Score 357; DB 9; Length 350;
Best Local Similarity 28.9%; Pred. No. 5.5e-25;
Matches 111; Conservative 59; Mismatches 144; Indels 70; Gaps 12;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSG----GPVFFP 56
           || : :||  | |  | |  |:|   :|| |||   ||||:|: ::: :|    | :
Db       1 MKSVRFYGIRDTRV-EDVDVPKILEKDDVIIKVKVAGICGSDISKYSKTGPHMVGEIL-- 57

Qy      57 KQGTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVA 116
                          ||||||  | :||  | | | ||||||     | :
Db      58 ----------------GHEFSGEVAQVGKEVRSFKIGDRVAVCPAMPCFE--------- 91

Qy     117 QDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEP 176
                 |  |: |  : | :::  |     | ||||    | :::|:|| |  :  | :||
Db      92 -----CDECKKGLYSRCNNVAIIGNKELGGCFAEYTKVKERNLIKIPDEISYETAAALEP 146

Qy     177 ISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELG 236
           : :| | : |:  : | | :||| ||||| :|   :   : ||:  :    :  |||||
Db     147 VCIAGHGLFRSEAKVGDTVVVLGTGPIGLFSIQWAKIFGSTKIIAVDVFDEKLDLAKELG 206

Qy     237 AEVFDPSTCDDANAVLKAMVPE------NEGFHAAFDCSGVPQTFTTSIVATGPSGIAVN 290
           |::     |   ||  | :| |       :|     : :| | |    ::      |  :
Db     207 ADI-----C--INAKEKNIVEEIKRLTDGDGADIVIESAGTPLTCGQVLLLAKKGGTVLY 259

Qy     291 VAVWGDHPIGFMPMSLTYQEKYATGSMCYTVK------DFQEVVKALEDGLISLDK---- 340
             |    | | : ::    ||     :  |||      |    |    ||  : |
Db     260 AGV----PYGDVALTREQFEKIVRSEL--TVKGTWFGNSFPFPGKEWSAGLYHMQKGDMN 313

Qy     341 ARKMITGKVHLKDG---VEKGFKQ 361
             |::| :::|::     || :|:
Db     314 VEKLVTHRINLEEAPAYFEKVYKR 337
```

RESULT 6
US-09-908-744-50

; Sequence 50, Application US/09908744
; Publication No. US20030068791A1
; GENERAL INFORMATION:
; APPLICANT: Miasnikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelainen Mira
; APPLICANT: Gros H+kan
; APPLICANT: Toivari Mervi
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koivuranta Kari
; APPLICANT: Londesborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Menut Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/FI01/00051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03
; PRIOR APPLICATION NUMBER: US 08/110,672
; PRIOR FILING DATE: 1993-08-24
; PRIOR APPLICATION NUMBER: US 07/973,325
; PRIOR FILING DATE: 1992-11-05
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 50
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-908-744-50

Query Match 17.5%; Score 352.5; DB 9; Length 354;
Best Local Similarity 27.6%; Pred. No. 1.5e-24;
Matches 108; Conservative 56; Mismatches 176; Indels 51; Gaps 9;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || |  ||  |:|: || | | |:: :|: |||   ||||:||  :
Db       1 MKALNLYGIQDLRFEET-PAPSIEHDDDIIIKVKAVGICGSDLSRY-------------- 45

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             |:  |   :  ||||:| ||::|  ||    |||||   |  |                |
Db      46 -KKLGPYVPGMTFGHEFAGEVVKIGRSVTGFSIGDRVAACPTYTC-------------G 90

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           |  || | |  |  ||  |      | :||||     |:: ||: :  |  ||:|| ||
Db      91 QCRYCQLGEPTRCERLSVIG-ARHPGAYAEYVKLPAKHVIPLPNVVNYDEAALIEPASVV 149

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
           |   |  :|| :  ::| | ||| :   :   |  :   :   | :|||:|
Db     150 AHGFYRTNIKPGASVAIMGVGSIGLLAVQWAKIFGATTVFAIDIDEQKLNVANQLGADVL 209

Qy     241 DPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIG 300
           |    |:  :        |   | : :| | |           |  | : :    |
Db     210 ISSLQRPAHKQILEYT-NGIGVDVAVESAGTPSTSAQVFALPKKGGEVVFLGI----PYA 264

Qy     301 FMPMSLTYQEKYATG----------SMCYTVKDFQEVVKALEDGLISLDKARKMITGKV 349
           : :   | ||                |  :  |::   :  :  | :::     ||: ::
Db     265 DVQIERFYFEKIVRNELHVYGSWNALSSPFPGKEWATTIHYMSSGQLNV---APMISYRL 321

Qy     350 HLKDGVEKGFKQLIEHKENNVKILVTPNEVS 380
           |  | |  |:|: : :    |:|  | ::|
Db     322 PLAKGPET-FQQIAKGELKPTKVLFYPEKLS 351
```

RESULT 7
US-09-815-242-10187
; Sequence 10187, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10187
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10187

Query Match 17.0%; Score 342; DB 10; Length 358;
Best Local Similarity 28.2%; Pred. No. 1.4e-23;
Matches 110; Conservative 66; Mismatches 162; Indels 52; Gaps 16;

```
Qy       1 MKGLLYYGTNDIRYSE-TVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQG 59
           || |  :|      |    ||:| :  | || |::    ||| |:| :    |
Db       1 MKALARFGKAFGGYKMIDVPQP-MCGPEDVVIEIKAAAICGADMKHYNVDSG-------- 51

Qy      60 TKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDL 119
             |: :        ||||:| : :||  |   | | ||  : : |
Db      52 -SDEFNSIR-----GHEFAGCIAQVGEKVKDWKVGQRVVSDNSGHV----------C 92

Qy     120 GLCMACQSGSPNCCASLSFCGLGGAS--GGFAEY-VVYGE------DHMVKLPDSIPDDI 170
           |:| ||: |   ||       ||   :  |||::| :| ||         : ::|| :  :
Db      93 GVCPACEQGDFLCCTEKVNLGLDNNTWGGGFSKYCLVPGEILKIHRHALWEIPDGVDYED 152

Qy     171 GALVEPISVAWHAV-ERARFQPGQTALVLGGGPIGLATILALQGHHAGKIV---CSEPAL 226
           |:::||  |: :: ::::| |||  :|:| ||:|| ::   :   |  ||     |
Db     153 AAVLDPICNAYKSIAQQSKFLPGQDVVVIGTGPLGLFSVQMARIMGAVNIVVVGLQEDVA 212

Qy     227 IRRQFAKELGA-EVFDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPS 285
           :|   ||||||  | : || |       :   :| |     :|||       :|    |:
Db     213 VRFPVAKELGATAVVNGSTEDVVARCQQICGKDNLGL--VIECSGANIALKQAIEMLRPN 270

Qy     286 GIAVNVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMI 345
           |  | | : |  |: |    :|   |   | | |    ::  :: |  | |   | : ||
Db     271 GEVVRVGM-GFKPLDFSINDITAWNKSIIGHMAYDSTSWRNAIRLLASGAI---KVKPMI 326

Qy     346 TGKVHLKDGVEKGFKQLIEHKENNVKILVT 375
           | :: |    | ||  :::  :  :|:::|
Db     327 THRIGLSQWRE-GFDAMVD--KTAIKVIMT 353
```

[redacted]ULT 8
US-09-908-744-70
; Sequence 70, Application US/09908744
; Publication No. US20030068791A1
; GENERAL INFORMATION:
; APPLICANT: Miasnikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelainen Mira
; APPLICANT: Gros H+kan
; APPLICANT: Toivari Mervi
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koivuranta Kari
; APPLICANT: Londesborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Menut Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/FI01/00051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03
; PRIOR APPLICATION NUMBER: US 08/110,672
; PRIOR FILING DATE: 1993-08-24
; PRIOR APPLICATION NUMBER: US 07/973,325
; PRIOR FILING DATE: 1992-11-05
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 70
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Bacillus halodurans (deduced sequence)
US-09-908-744-70

Query Match 16.8%; Score 337.5; DB 9; Length 343;
Best Local Similarity 28.7%; Pred. No. 3.5e-23;
Matches 111; Conservative 57; Mismatches 162; Indels 57; Gaps 19;

```
Qy       1 MKGLL--YYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQ 58
           || |:   :||       |  |||     .: ||||| | |:||:|:   || |    :|
Db       1 MKALVKTQHGTGHFAVQEK-PEP-TPGKHQVKIKVKYTGVCGSDI--HTYEG---HYP-- 51

Qy      59 GTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEAT-SHCSDRSRYKDTVAQ 117
                  |:  |||||| :||:| |||    ||||  | | | |
Db      52 --------VAAPVTLGHEFSGEIVELGEGVTGFNVGDRVTSETTYSIC---------- 91

Qy     118 DLGLCMACQSGSPNCCASLSFCGLGG-ASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEP 176
             | |  | ||  | |:     |||     | ||:||:  :: :  ||  : |   |: ||
Db      92 --GKCSYCTSGDYNLCSHRK--GLGNQQDGSFAKYVIARQESLHHLPAGVDDRSAAMTEP 147

Qy     177 ISVAWHAVERARFQPGQTALVLGGGPIGLATILALQ--GHHAGKIVCS--EPALIRRQFA 232
           ::   ||: :     |   :| | |||||   || |    | | :: :     :| : |
Db     148 LACTHHAIAKTSINKGDLVVVTGPGPIGL---LAAQVAKSHGGTVIITGLSNDQVRLKKA 204

Qy     233 KELGAEVFDPSTCDDANAVLKAMVPE-NEGFHA--AFDCSGVPQTFTTSIVATGPSGIAV 289
           ||:| :       |     :| :| |  :|: |    :|||         |       |
Db     205 KEVGIDY----AIDTQEVDIKELVSELTDGYGADVVLECSGAVPAAKQGIDLLRKKGQYA 260

Qy     290 NVAVWGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKV 349
             | ::    | |    :  :|    ||       |::  :  | :  ::   |: ::| :
Db     261 QVGLFAQPEIQFNFEKIIQKEISVVGSRSQKPADWEPALSLLNEKKVN---AKTLVTHEY 317

Qy     350 HLKDGVEKGFKQLIEHKENNVKILVTP 376
             : :  :| :  :       :|:|:||
Db     318 TISEW-DKAYHAI--KSGEAIKVLLTP 341
```

RESULT 9
US-09-908-744-51
; Sequence 51, Application US/09908744
; Publication No. US20030068791A1
; GENERAL INFORMATION:
; APPLICANT: Miasnikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelainen Mira
; APPLICANT: Gros H+kan
; APPLICANT: Toivari Mervi
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koivuranta Kari
; APPLICANT: Londesborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Menut Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/FI01/00051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03
; PRIOR APPLICATION NUMBER: US 08/110,672
; PRIOR FILING DATE: 1993-08-24
; PRIOR APPLICATION NUMBER: US 07/973,325
; PRIOR FILING DATE: 1992-11-05
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 51
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-908-744-51

Query Match 16.8%; Score 337.5; DB 9; Length 350;
Best Local Similarity 27.7%; Pred. No. 3.6e-23;
Matches 112; Conservative 67; Mismatches 144; Indels 81; Gaps 15;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIK--NPNDVKIKVSYCGICGTD----LKEFTYSGGPVF 54
           ||  : :||||:|: :     |||    ::|||||   ||||:|    || : |   ||
Db       1 MKAAVLHGTNDMRFEDI----EIKPCESDEVKIKVMAAGICGSDPPRVLKHWKY---PV- 52

Qy      55 FPKQGTKDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDT 114
                     |  ||||||| : |||   | :|| ||||       |::
Db      53 ---------------PAIPGHEFSGVIAEVGKDVKNVKVGDRVVAIPFIPCNE------- 90

Qy     115 VAQDLGLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALV 174
                 |  |: |  : |      |   : | |||||     ::: : |   :|  |::
Db      91 -------CEYCKRGLFSLCDDHGMLG-AKSFGAFAEYVNIKATNVLPIGDMDFED-AAMI 141

Qy     175 EPISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKE 234
           ||::|| | |      | | |  |:| | :|   |  |:   || |:  : :  : :||
Db     142 EPLAVAMHGVLNIGVQVGDTVAVMGSGTMGQLVIQGLKIAGAGTIIAVDISDNKLRESKE 201

Qy     235 LGAEVFDPSTCDDANAVLKAM-VPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
           |||::       | | | |   :   :|   | :|:|   |    :: |
Db     202 LGADII--INAKDINPVEKIKELTGGKGVDIALECAGSKITQEQCLLITKKKS------- 252

Qy     294 WGDHPIGFMPMS---LTYQEK--------------YATGSMCYTVKDFQEVVKALEDGL 335
               |||: ::   :|  |:                : : |  :  ::: : :  : :|
Db     253 ----KIGFLGIAYSDITLSEEAFENIFRKELELKGFWNSYSAPFPGQEWTKGINLVNEGK 308

Qy     336 ISLDKARKMITGKVHLKDGVEKGFKQLIEHKENNVKILVTPNEV 379
```

```
Db    309 I---KLKEMVSHRFSLED-TYKAFEMIRDRKEEFNKILILPQGV 348


RESULT 10
US-09-922-501-15
; Sequence 15, Application US/09922501
; Patent No. US20020120119A1
; GENERAL INFORMATION:
;  APPLICANT: Dartois, Veronique A.
;  APPLICANT:  Hoch, James A.
;  APPLICANT:  Valle, Fernando
;  APPLICANT:  Kumar, Manoj
;  TITLE OF INVENTION: 2, 5-DKG PERMEASES
;  FILE REFERENCE: P-SR 4877
;  CURRENT APPLICATION NUMBER: US/09/922,501
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US 09/633,294
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 09/677,032
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 343
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: environmental source
US-09-922-501-15

  Query Match             16.7%;  Score 336;  DB 10;  Length 343;
  Best Local Similarity   28.9%;  Pred. No. 4.8e-23;
  Matches  105;  Conservative   63;  Mismatches  149;  Indels   46;  Gaps   13;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          ||  ::  :   |:|  ::    |   | :|:|:::  ||||:||  : : |    |
Db      1 MKAIVIHQAKDLRVEDSA--VEAPGPGEVEIRLAAGGICGSDLHYYNHGG----FGTVRL 54

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
          |:       |:  |||| || |  :| ||: :  || |||  :  |               |
Db     55 KE-------PMILGHEVSGHVAALGEGVSGLAIGDLVAVSPSRPC-------------G 93

Qy    121 LCMACQSGSPNCCASLSFCG----LGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEP 176
          |  |  |  | | :: | |          | | | :|      ||  : :     |: ||
Db     94 ACDYCLKGLANHCFNMRFYGSAMPFPHIQGAFRERLVAKASQCVK-AEGLSAGEAAMAEP 152

Qy    177 ISVAWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELG 236
          :||  ||. ||    |:  || | ||||  :||| :   | :|| :: :     ||: :|
Db    153 LSVTLHATRRAGEMLGKRVLVTGCGPIGTLSILAARRAGAAEIVAADLSERALGFARAVG 212

Qy    237 AEVFDPSTCDDANAVLKAMVP--ENEG-FHAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
          |:    :  :| :     :||  ||:| |   ::|||         | |  | |: | : :
Db    213 AD-RTVNLSEDRD----GLVPFSENKGYFDVLYECSGAQPALVAGIQALRPRGVIVQLGL 267

Qy    294 WGDHPIGFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKD 353
          |:  :    |::| :|    ||  :  ::|   || :: ||| :   : :||  : | |
Db    268 GGE--MSLPMMAITAKELDLRGSFRFH-EEFAVAVKLMQGGLIDV---KPLITHTLPLAD 321

Qy    354 GVE 356
          ::
Db    322 ALQ 324

RESULT 11
US-09-908-744-52
; Sequence 52, Application US/09908744
; Patent No. US20030068791A1
; GENERAL INFORMATION:
;  APPLICANT: Miasnikov Andrei
;  APPLICANT:  Ojamo Heikki
;  APPLICANT:  Povelainen Mira
;  APPLICANT:  Gros H+kan
;  APPLICANT:  Toivari Mervi
;  APPLICANT:  Richard Peter
;  APPLICANT:  Ruohonen Laura
;  APPLICANT:  Koivuranta Kari
;  APPLICANT:  Londesborough John
;  APPLICANT:  Aristidou Aristos
;  APPLICANT:  Penttil, Merja
;  APPLICANT:  Plazanet-Menut Claire
;  APPLICANT:  Deutscher Josef
;  TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
;  FILE REFERENCE: 1427.0010005
;  CURRENT APPLICATION NUMBER: US/09/908,744
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: PCT/FI01/00051
;  PRIOR FILING DATE: 2001-01-22
;  PRIOR APPLICATION NUMBER: US 09/488,581
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: US 08/790,585
;  PRIOR FILING DATE: 1997-01-29
;  PRIOR APPLICATION NUMBER: US 08/368,395
;  PRIOR FILING DATE: 1995-01-03
;  PRIOR APPLICATION NUMBER: US 08/110,672
;  PRIOR FILING DATE: 1993-08-24
;  PRIOR APPLICATION NUMBER: US 07/973,325
;  PRIOR FILING DATE: 1992-11-05
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 52
;   LENGTH: 352
;   TYPE: PRT
;   ORGANISM: Clostridium difficile
US-09-908-744-52

  Query Match             16.6%;  Score 334;  DB 9;  Length 352;
  Best Local Similarity   28.5%;  Pred. No. 7.6e-23;
  Matches  110;  Conservative   61;  Mismatches  163;  Indels   52;  Gaps   12;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          |:  : |   |:|| | |  ||| :   | : | | ||||:||     ||
Db      5 MRASVLYNVGDVRY-EMVDIPEITD-TQVLVNVKYVGICGSDLPRSMVSG---------- 52

Qy     61 KDKISG-YELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDL 119
             :||  : ||  |||||| ||::|  |  :  ||||||     |
Db     53 ---LSGNTKYPLILGHEFSGEVVKIGEKVKHINVGDRVAVAPLVPC---------- 95

Qy    120 GLCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISV 179
          | |  |:   |   :  |   :| |||||   |:|::||||::  :  | :|| ::
Db     96 GKCDYCNEGNFGLCDDYNIIGT-RVNGAFAEYVRVPEEHILKLPDTLDYETAAGIEPATI 154

Qy    180 AWHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEV 239
          |:| : ::  : | : :|||  ||||   |  :  | ||:  :     : : :| |||
Db    155 AYHGISKSNIRVGDSVVVLGCGPIGQFVIQWAKVFGASKIIAVDIFDEKLELSKLLGANY 214

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
          |   : | : :     | |     : :|   |   |:      |  | | :   |
Db    215 ILNS--KEVNVIKEIKKITNGGADVVIETAGSRFTQEQSLFVAKKRGNIVFVGI--SHT- 269

Qy    300 GFMPMS----------LTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGK 348
          :|:|           || :  : : :  | : :   :  :| | |     : ||: |
Db    270 -ELPLSADATECILRGELTLKGSWNSYTSPYPGRAWTATLDFMEKGDIIF---KPMISDK 325

Qy    349 VHLKDGVEKGFKQLIEHKENNVKILV 374
          : | : |     :: : : |  ||||
Db    326 IGLNE-VGDFLSKMSKREINFNKILV 350


RESULT 12
US-09-925-300-1326
; Sequence 1326, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
```

```
Qy    241 DPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIG 300

Qy    301 FMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 FMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGFK 360

Qy    361 QLIEHKENNVKILVTPNEVS 380
          ||||||||||||||||||||
Db    361 QLIEHKENNVKILVTPNEVS 380


RESULT 2
US-10-147-003-2
; Sequence 2, Application US/10147003
; Publication No. US20030032153A1
; GENERAL INFORMATION:
;  APPLICANT: Yamamoto, Hiroaki
;  APPLICANT:  Kimoto, No. US20030032153A1Ihiro
;  TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE,
;  TITLE OF INVENTION:  METHODS FOR PRODUCING SAME, AND METHODS FOR PRODUCING
;  TITLE OF INVENTION:  OPTICALLY ACTIVE ALCOHOL USING THE DEHYDROGENASE
;  FILE REFERENCE: 06501-109001
;  CURRENT APPLICATION NUMBER: US/10/147,003
;  CURRENT FILING DATE:  2002-05-16
;  PRIOR APPLICATION NUMBER: JP 2001-159647
;  PRIOR FILING DATE: 2001-05-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 385
;   TYPE: PRT
;   ORGANISM: Kluyveromyces lactis
US-10-147-003-2

  Query Match           52.9%;  Score 1065.5;  DB 9;  Length 385;
  Best Local Similarity  55.0%;  Pred. No. 9.6e-91;
  Matches  208;  Conservative   56;  Mismatches  111;  Indels    3;  Gaps    2;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          |: | |:|  ||||:: : || |:  : ::|:||:|||||:|| |  |  ||:|||: |
Db      1 MRALAYFGKQDIRYTKDLEEPVIETDDGIEIEVSWCGICGSDLHE--YLDGPIFFPEDGK 58

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             :|| ||   ||| || | :||  ||::| || | ||||  | |   : :
Db     59 VHDVSGLGLPQAMGHEMSGIVSKVGPKVTNIKAGDHVVVEATGTCLDHYTWPNAAHAKDA 118

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
          | ||| |  |||| | | |||  |||||| ||  | |:||:|:::| |: ||||||||:
Db    119 ECAACQRGFYNCCAHLGFMGLGVHSGGFAEKVVVSEKHVVKIPNTLPLDVAALVEPISVS 178

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
          ||||  :: | ||:||||| |||||||||||||| | ||| |||| |||  | :|| | |
Db    179 WHAVRISKLQKGQSALVLGAGPIGLATILALQGHGASKIVVSEPAEIRRNQAAKLGVETF 238

Qy    241 DPST-CDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
          |||   :|| :|| : |  |||  |:|||||  || | : ||   |: ||:|:||  ||
Db    239 DPSEHKEDAVNILKKLAPGGEGFDFAYDCSGVKPTFDTGVHATTFRGMYVNIAIWGHKPI 298

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
          | || :| |||: |||||||:|||::||:|| :| |::|||| :|||:  ::||  |||
Db    299 DFKPMDVTLQEKFVTGSMCYTIKDFEDVVQALGNGSIAIDKARHLITGRQKIEDGFTKGF 358

Qy    360 KQLIEHKENNVKILVTPN 377
          :|: ||| |:|||:|||
Db    359 DELMNHKEKNIKILLTPN 376


RESULT 3
US-09-815-242-11979
; Sequence 11979, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11979
;   LENGTH: 363
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11979

  Query Match           28.2%;  Score 567.5;  DB 10;  Length 363;
  Best Local Similarity  36.1%;  Pred. No. 1.6e-44;
  Matches  136;  Conservative   73;  Mismatches  141;  Indels   27;  Gaps   11;

Qy      1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
          |:  :::| :|||  | || |.   |  |:|:| :|||||:|| |  |  |||| | :
Db     11 MRAAVWHGRHDIRV-EDVPLPAEPPPGWVQIRVHWCGICGSDLHE--YLAGPVFIPVE-A 66

Qy     61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
             ::| :     ||||||| :| :|:|||    |  || :|  ||              |
Db     67 PHPLTGLKDQCILGHEFSGEIVRLGNGVTGFAVGQAVAADACQHC-------------G 112

Qy    121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
          |  |: |  | | :|:| ||   :| |||||    : :  ||   | : |||:||::|
Db    113 TCYYCRHGLYNICENLAFTGLMN-NGAFAEYVNVPANLLYALPAGFPSEAGALIEPLAVG 171

Qy    181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGA-EV 239
          |||::|    ||  :|:| | |||:||:  :   | :::  | :  |:  | |:|| :|
Db    172 MHAVKKAGSLLGQNVVVVGAGTIGLSTIMCARAAGAAQVIALEMSSARKAKALEVGASQV 231

Qy    240 FDPSTCDDANAVLKAMVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPI 299
          |||  | ||   ::|:     |   :|:| |   |   :| |   :|  | | :: : |
Db    232 LDPSRC-DALGEIRALT-GGLGADVSFECIGNKHTAKLAIDAIRKAGKCVLVGIF-EEPS 288

Qy    300 GFMPMSLTYQEKYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVHLKDGVEKGF 359
          |    |   ||   |:: |   :| :|:  : || : :     ::||:: |:: ||:||
Db    289 EFNFFELVSTEKQLLGALAYN-GEFADVIAFIADGRLDI---APLVTGRIGLEEIVERGF 344

Qy    360 KQLIEHKENNVKILVTP 376
          ::|: :||:||||:|:|
Db    345 EELVNNKEHNVKIIVSP 361


RESULT 4
US-09-815-242-11975
; Sequence 11975, Application US/09815242
; Patent No. US20020061569A1
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 2, 2003, 12:51:13 ; Search time 1361 Seconds
(without alignments)
24.092 Million cell updates/sec

Title: US-10-020-674-2
Perfect score: 2013
Sequence: 1 MKGLLYYGTNDIRYSETVPE.....QLIEHKENNVKILVTPNEVS 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2013	100.0	380	9	US-10-020-674-2	Sequence 2, Appli
2	1065.5	52.9	385	9	US-10-147-003-2	Sequence 2, Appli
3	567.5	28.2	363	10	US-09-815-242-11979	Sequence 11979, A
4	455.5	22.6	352	10	US-09-815-242-11975	Sequence 11975, A
5	357	17.7	350	9	US-09-908-744-53	Sequence 53, Appl
6	352.5	17.5	354	9	US-09-908-744-50	Sequence 50, Appl
7	342	17.0	358	10	US-09-815-242-10187	Sequence 10187, A
8	337.5	16.8	343	9	US-09-908-744-70	Sequence 70, Appl
9	337.5	16.8	350	9	US-09-908-744-51	Sequence 51, Appl
10	336	16.7	343	10	US-09-922-501-15	Sequence 15, Appl
11	334	16.6	352	9	US-09-908-744-52	Sequence 52, Appl
12	326.5	16.2	356	10	US-09-925-300-1326	Sequence 1326, Ap
13	319	15.8	343	10	US-09-815-242-5705	Sequence 5705, Ap
14	319	15.8	354	10	US-09-815-242-12464	Sequence 12464, A
15	315	15.6	347	10	US-09-815-242-13773	Sequence 13773, A
16	314	15.6	352	9	US-09-908-744-69	Sequence 69, Appl
17	308	15.3	341	10	US-09-815-242-13833	Sequence 13833, A
18	307.5	15.3	346	10	US-09-741-669-421	Sequence 421, App
19	307.5	15.3	346	10	US-09-815-242-10206	Sequence 10206, A
20	306	15.2	342	10	US-09-815-242-10958	Sequence 10958, A
21	284.5	14.1	457	9	US-10-002-245-4	Sequence 4, Appli
22	282.5	14.0	455	10	US-09-775-009-7	Sequence 7, Appli
23	277	13.8	347	10	US-09-815-242-12466	Sequence 12466, A
24	268.5	13.3	305	10	US-09-815-242-5706	Sequence 5706, Ap
25	265.5	13.2	343	10	US-09-741-669-412	Sequence 412, App
26	264	13.1	340	10	US-09-971-361-10	Sequence 10, Appl
27	259	12.9	354	9	US-10-166-087-6	Sequence 6, Appli
28	255	12.7	369	10	US-09-854-122-40	Sequence 40, Appl
29	252.5	12.5	359	10	US-09-815-242-13653	Sequence 13653, A
30	246	12.2	346	10	US-09-815-242-10162	Sequence 10162, A
31	239.5	11.9	336	10	US-09-815-242-12430	Sequence 12430, A
32	239.5	11.9	339	10	US-09-815-242-5496	Sequence 5496, Ap
33	236	11.7	352	10	US-09-954-314-18	Sequence 18, Appl
34	231	11.5	336	10	US-09-815-242-13832	Sequence 13832, A
35	230.5	11.5	349	9	US-09-908-744-49	Sequence 49, Appl
36	228.5	11.4	342	10	US-09-815-242-12104	Sequence 12104, A
37	228	11.3	368	9	US-09-738-626-3857	Sequence 3857, Ap
38	225.5	11.2	375	9	US-09-981-353-113	Sequence 113, App
39	225.5	11.2	386	9	US-09-854-133-81	Sequence 81, Appl
40	225.5	11.2	386	10	US-09-738-973-81	Sequence 81, Appl
41	222	11.0	357	10	US-09-947-027-9	Sequence 9, Appli
42	222	11.0	357	12	US-10-091-009-9	Sequence 9, Appli
43	220.5	11.0	264	10	US-09-815-242-11710	Sequence 11710, A
44	220	10.9	362	10	US-09-947-150-2	Sequence 2, Appli
45	220	10.9	362	10	US-09-947-027-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-020-674-2
; Sequence 2, Application US/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Onodera, Keiko
; APPLICANT: Tani, Yoshiki
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
; FILE REFERENCE: 06501-092001
; CURRENT APPLICATION NUMBER: US/10/020,674
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: JP 2000-333363
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Pichia angusta
US-10-020-674-2

Query Match 100.0%; Score 2013; DB 9; Length 380;
Best Local Similarity 100.0%; Pred. No. 8.9e-179;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
```

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1326
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1326

Query Match 16.2%; Score 326.5; DB 10; Length 356;
Best Local Similarity 30.1%; Pred. No. 3.9e-22;
Matches 86; Conservative 44; Mismatches 119; Indels 37; Gaps 6;

```
Qy      14 YSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCP 73
           ::|  | ||   ||:| :::   ||||:|:  : |     |  |:           |:
Db      20 FAENYPIPE-PGPNEVLLRMHSVGICGSDVHYWEYGRIGNFIVKK----------PMVL 67

Qy      74 GHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLGLCMACQSGSPNCC 133
           ||| |||| :||| |  :|||||||:|  :      |  |           |: |  |
Db      68 GHEASGTVEKVGSSVKHLKPGDRVAIEPGA-----PRENDEF---------CKMGRYNLS 113

Qy     134 ASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQ 193
            |: ||      |    :  :       ||||::  : |||:||:||  ||  |     |
Db     114 PSIFFCATPPDDGNLCRFYKHNAAFCYKLPDNVTFEEGALIEPLSVGIHACRRGGVTLGH 173

Qy     194 TALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVFDPSTCDDANAVLK 253
           || | ||||: |:|  :   | ::| :: :  |   |||:||::    : :     : :
Db     174 KVLVCGAGPIGMVTLLVAKAMGAAQVVVTDLSATRLSKAKEIGADLVLQISKESPQEIAR 233

Qy     254 AM------VPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAV 293
           :        ||        :|:|   :  ' | ||   |  | | :
Db     234 KVEGQLGCKPE-----VTIECTGAEASIQAGIYATRSGGTLVLVGL 274
```

RESULT 13
US-09-815-242-5705
; Sequence 5705, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5705
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5705

Query Match 15.8%; Score 319; DB 10; Length 343;
Best Local Similarity 27.7%; Pred. No. 1.8e-21;
Matches 100; Conservative 54; Mismatches 153; Indels 54; Gaps 14;

```
Qy      20 EPEIKNPND--VKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEF 77
           : |:  | |  ||||| | ||||||:   || |            |::    |: ||||
Db      18 DKEVATPLDDKVKIKVHYAGICGTDI--HTYEG----------HYKVN---FPVTLGHEF 62

Qy      78 SGTVVEVGSGVTSVKPGDRVAVEATSH-CSDRSRYKDTVAQDLGLCMACQSGSPNCCASL 136
           || :|||| |   | ||||  | | : |::              |  |:|   | |
Db      63 SGEIVEVGADVKDFKVGDRVTSETTFYVCNE-------------CEYCESKDYNLCNHR 108

Qy     137 SFCGLG-GASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTA 195
           |:|    | |  ||:  |: :  :|| :     |: ||:: | | | : :   | |
Db     109 K--GIGTQVDGAFTNYVIAREESLHHIPDEVSYQSAAMTEPLACAHHGVSKIQVNSGDVA 166

Qy     196 LVLGGGPIGLATILALQGHHAGKIVCS-EPALIRRQFAKELGAEVFDPSTCDDANAVLKA 254
           :|:| |||||     |:   |  :|   :   :|   |: | :         |    :
Db     167 VVMGPGPIGLLVAQVLKSKGATVVVTGLDNDKVRLDKAEALHMDYVVNLQQTDLKTYING 226

Qy     255 MVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTYQEKYAT 314
           : :  |     :|||         :       |    : :: |  | |    :  :|
Db     227 -ITDGYGADVVVECSGAVPAARQGLDILRKKGFYSQIGIFKDAEIPFDMEKVIQKEITVV 285

Qy     315 GSMCYTVKDFQEVVKALEDGLISLDKARKMITGKV----------HLKDGVEKGFKQLI 363
           ||       |::  :: : |||::   |  ::| |:           ||| |  :| | |:
Db     286 GSRSQKPADWEPSLQLMADGLVN---AEALVT-KIYDISKWDEAYQHLKSG--EGIKALL 339

Qy     364 E 364
           :
Db     340 K 340
```

RESULT 14
US-09-815-242-12464
; Sequence 12464, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12464
;   LENGTH: 354
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12464

  Query Match             15.8%;  Score 319;  DB 10;  Length 354;
  Best Local Similarity   27.7%;  Pred. No. 1.9e-21;
  Matches  100;  Conservative   54;  Mismatches  153;  Indels   54;  Gaps   14;

Qy      20 EPEIKNPND--VKIKVSYCGICGTDLKEFTYSGGPVFFPKQGTKDKISGYELPLCPGHEF 77
           : |:  | |  |||||  | ||||||:   || |           |::    |:  ||||
Db      21 DKEVATPLDDKVKIKVHYAGICGTDI--HTYEG----------HYKVN---FPVTLGHEF 65

Qy      78 SGTVVEVGSGVTSVKPGDRVAVEATSH-CSDRSRYKDTVAQDLGLCMACQSGSPNCCASL 136
           || :||||: |   | ||||  | | : |::             |  |:|   | |
Db      66 SGEIVEVGADVKDFKVGDRVTSETTFYVCNE-------------CEYCESKDYNLCNHR 111

Qy     137 SFCGLG-GASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVAWHAVERARFQPGQTA 195
             |:|    | |  ||:  |: :  :|| :     |: ||:: | | | : :   |  |
Db     112 K--GIGTQVDGAFTNYVIAREESLHHIPDEVSYQSAAMTEPLACAHHGVSKIQVNSGDVA 169

Qy     196 LVLGGGPIGLATILALQGHHAGKIVCS-EPALIRRQFAKELGAEVFDPSTCDDANAVLKA 254
           :|:| |||||     |:   |  :|   :   :|   |: |  :          |    :
Db     170 VVMGPGPIGLLVAQVLKSKGATVVVTGLDNDKVRLDKAEALHMDYVVNLQQTDLKTYING 229

Qy     255 MVPENEGFHAAFDCSGVPQTFTTSIVATGPSGIAVNVAVWGDHPIGFMPMSLTYQEKYAT 314
           : :  |     :|||         :       |    : :: |  | |    :  :|
Db     230 -ITDGYGADVVVECSGAVPAARQGLDILRKKGFYSQIGIFKDAEIPFDMEKVIQKEITVV 288

Qy     315 GSMCYTVKDFQEVVKALEDGLISLDKARKMITGKV----------HLKDGVEKGFKQLI 363
           .||      |::  :: : |||::   |  ::| |:          ||| |  :| | |:
Db     289 GSRSQKPADWEPSLQLMADGLVN---AEALVT-KIYDISKWDEAYQHLKSG--EGIKALL 342

Qy     364 E 364
           :
Db     343 K 343


RESULT 15
US-09-815-242-13773
; Sequence 13773, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16

;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13773
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Salmonella typhi
US-09-815-242-13773

  Query Match             15.6%;  Score 315;  DB 10;  Length 347;
  Best Local Similarity   28.2%;  Pred. No. 4.3e-21;
  Matches   81;  Conservative   48;  Mismatches  120;  Indels   38;  Gaps    6;

Qy       1 MKGLLYYGTNDIRYSETVPEPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGPVFFPKQGT 60
           || :: :   |:| |  | |:::  :|| :||   |:||:|:      |   :
Db       1 MKSVVIHAEGDVRVEER-PLPQLQAEDDVLVKVVSSGLCGSDIPRIFAQGAHYY------ 53

Qy      61 KDKISGYELPLCPGHEFSGTVVEVGSGVTSVKPGDRVAVEATSHCSDRSRYKDTVAQDLG 120
                 |:  |||||| |   |:||| ::||| ||      |
Db      54 ---------PITLGHEFSGYVESYGTGVTDMQPGDAVACVPLLPCFH------------- 91

Qy     121 LCMACQSGSPNCCASLSFCGLGGASGGFAEYVVYGEDHMVKLPDSIPDDIGALVEPISVA 180
           |  |: |  : |    | |   : || |||||    :: :||  :| : || :|||:|
Db      92 -CPQCERGYFSLCKQYQFVG-SRSEGGNAEYVVVKRANLFRLPSDMPIEDGAFIEPITVG 149

Qy     181 WHAVERARFQPGQTALVLGGGPIGLATILALQGHHAGKIVCSEPALIRRQFAKELGAEVF 240
           ||   |:   |:  :::| | |||  :   :   |  :    :  : || |||
Db     150 LHAFHLAQGCEGKNVIIVGAGTIGLLALQCARELGARSVTAIDINPQKLELAKALGA--- 206

Qy     241 DPSTCDDANAV---LKAMVPENEGFHAAFDCSGVPQTFTTSIVATGP 284
           ||:        ::  : : :      : :| ||| : :|  |||
Db     207 -THTCNSREMTADDIQTALSDIQFDQLVLETAGTPQTVSLAIDITGP 252


Search completed: May  2, 2003, 15:32:41
Job time : 1363 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 07:40:56 ; Search time 3103 Seconds
(without alignments)
10720.112 Million cell updates/sec

Title: US-10-020-674-1
Perfect score: 1143
Sequence: 1 atgaaaggtttactttatta.....cgccgaacgaggtttcctaa 1143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		274	24.0	103682	8	SCU12980	U12980 Saccharomyc
2	c	163	14.3	915	11	CNS06I04	AL399674 T3 end of
3		83	7.3	1041	6	AX366887	AX366887 Sequence
4	c	83	7.3	16585	1	AB007638	AB007638 Bacillus
5	c	83	7.3	213190	1	BSUB0004	Z99107 Bacillus su
6		69.8	6.1	339	6	AX312470	AX312470 Sequence
7	c	67.4	5.9	717	11	CNS06K6L	AL402499 T7 end of
8		67	5.9	1126	8	AY133848	AY133848 Arabidops
9		67	5.9	1269	8	AF370161	AF370161 Arabidops
10		67	5.9	1347	8	AY085213	AY085213 Arabidops
11		65.8	5.8	1457	8	AB025969	AB025969 Prunus pe
12		64.4	5.6	10582	1	AE008211	AE008211 Agrobacte
13	c	64.4	5.6	11742	1	AE009402	AE009402 Agrobacte
14		64.2	5.6	81347	8	AB015478	AB015478 Arabidops
15		63.8	5.6	15088	1	AE004831	AE004831 Pseudomon
16		62.8	5.5	1404	3	AY052067	AY052067 Drosophil
17		62.2	5.4	1380	3	AY058731	AY058731 Drosophil
18		61.4	5.4	1456	8	AY037946	AY037946 Prunus ce
19		60.8	5.3	11248	1	AE006323	AE006323 Lactococc
20		59.6	5.2	1526	8	SPTMS1	X74422 S.pombe tms
21	c	59.6	5.2	40438	8	SPU41410	U41410 Schizosacch
22	c	59.6	5.2	40810	8	SPBC1773	AL033389 S.pombe c
23		59.6	5.2	305153	1	AP001520	AP001520 Bacillus
24		58.6	5.1	1572	8	AB042810	AB042810 Eriobotry
25		58.6	5.1	25025	3	CER04B5	Z70782 Caenorhabdi
26		57.4	5.0	933	6	AR165148	AR165148 Sequence
27		57	5.0	1396	8	AF323504	AF323504 Malus x d
28		57	5.0	1478	8	AB016256	AB016256 Malus dom
29		56.6	5.0	249050	1	AL596165	AL596165 Listeria
30	c	56.6	5.0	332635	1	AP003005	AP003005 Mesorhizo
31		56.6	5.0	349980	6	AX417038	AX417038 Sequence
32		56.6	5.0	349980	6	AX417041	AX417041 Sequence
33	c	55.8	4.9	24336	1	AE008841	AE008841 Salmonell
34		55.4	4.8	1038	6	AX377790	AX377790 Sequence
35		52.8	4.6	2921	3	DMAF002212	AF002212 Drosophil
36		52.8	4.6	72847	3	AC004266	AC004266 Drosophil
37	c	52.8	4.6	110000	3	AE001572_0	AE001572 Drosophil
38	c	52.8	4.6	152108	2	AC020282	AC020282 Drosophil
39		52.8	4.6	165761	3	AC095016	AC095016 Drosophil
40		52.8	4.6	174140	3	AC008225	AC008225 Drosophil
41		52.8	4.6	309357	3	AE003673	AE003673 Drosophil
42		52.2	4.6	1808	9	HSU07361	U07361 Homo sapien
43		52.2	4.6	2487	9	BC025295	BC025295 Homo sapi
44		52.2	4.6	2519	6	AX330493	AX330493 Sequence
45		52.2	4.6	2519	6	AX337243	AX337243 Sequence

ALIGNMENTS

RESULT 1
SCU12980
LOCUS SCU12980 103682 bp DNA linear PLN 05-MAR-1998
DEFINITION Saccharomyces cerevisiae chromosome I left arm sequence.
ACCESSION U12980 U00091
VERSION U12980.1 GI:2911250
KEYWORDS S.cerevisiae Ycr28p homolog; FLO9; GDH3; two alcohol/sorbitol dehydrogenase homologs; SIM1; CNE1; ACS1; S.pombe SPAC 1F7.03 homolog; S.cerevisiae Pip2p-like transcription factor homolog; GCV3, glycine cleavage H protein; PTA1; FUN9 transcript, essential gene, S.pombe SPAC 24B11.08c homolog; CDC24; CLN3/WHI1/DAF1; CYC3; PYK1; S.cerevisiae ORF 06283 homolog; Xenopus laevis GTP-binding protein DRG homolog; FUN12 transcript, essential gene; FUN19 transcript, S.cerevisiae orf 06265 homolog; FUN53 transcript, essential gene; FUN20 transcript, essential gene, S.pombe SPAC

8A4.06 homolog; SNC1; MYO2; S.cerevisiae ORF 06159 homolog; DRS2; MAK16; LTE1.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (sites)
AUTHORS Nagasu,T. and Hall,B.D.
TITLE Nucleotide sequence of the GDH gene coding for the NADP-specific glutamate dehydrogenase of Saccharomyces cerevisiae
JOURNAL Gene 37 (1-3), 247-253 (1985)
MEDLINE 86031359
PUBMED 2932370
REFERENCE 2 (sites)
AUTHORS Dumont,M.E., Ernst,J.F., Hampsey,D.M. and Sherman,F.
TITLE Identification and sequence of the gene encoding cytochrome c heme lyase in the yeast Saccharomyces cerevisiae
JOURNAL EMBO J. 6 (1), 235-241 (1987)
MEDLINE 87218469
PUBMED 3034577
REFERENCE 3 (sites)
AUTHORS Miyamoto,S., Ohya,Y., Ohsumi,Y. and Anraku,Y.
TITLE Nucleotide sequence of the CLS4 (CDC24) gene of Saccharomyces cerevisiae
JOURNAL Gene 54 (1), 125-132 (1987)
MEDLINE 87277425
PUBMED 3301539
REFERENCE 4 (sites)
AUTHORS Wickner,R.B.
TITLE Host function of MAK16: G1 arrest by a mak16 mutant of Saccharomyces cerevisiae
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (16), 6007-6011 (1988)
MEDLINE 88320371
PUBMED 3045810
REFERENCE 5 (sites)
AUTHORS Nash,R., Tokiwa,G., Anand,S., Erickson,K. and Futcher,A.B.
TITLE The WHI1+ gene of Saccharomyces cerevisiae tethers cell division to cell size and is a cyclin homolog
JOURNAL EMBO J. 7 (13), 4335-4346 (1988)
MEDLINE 89210821
PUBMED 2907481
REFERENCE 6 (sites)
AUTHORS McNally,T., Purvis,I.J., Fothergill-Gilmore,L.A. and Brown,A.J.
TITLE The yeast pyruvate kinase gene does not contain a string of non-preferred codons: revised nucleotide sequence
JOURNAL FEBS Lett. 247 (2), 312-316 (1989)
MEDLINE 89232143
PUBMED 2653861
REFERENCE 7 (sites)
AUTHORS Harris,S.D., Cheng,J., Pugh,T.A. and Pringle,J.R.
TITLE Molecular analysis of Saccharomyces cerevisiae chromosome I. On the number of genes and the identification of essential genes using temperature-sensitive-lethal mutations
JOURNAL J. Mol. Biol. 225 (1), 53-65 (1992)
MEDLINE 92260538
PUBMED 1583694
REFERENCE 8 (sites)
AUTHORS Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L., Alexandraki,D., Antoine,G., Anwar,R., Ballesta,J.P., Benit,P. et al.
TITLE The complete DNA sequence of yeast chromosome III
JOURNAL Nature 357 (6373), 38-46 (1992)
MEDLINE 92244356
PUBMED 1574125
REFERENCE 9 (sites)
AUTHORS Gerst,J.E., Rodgers,L., Riggs,M. and Wigler,M.
TITLE SNC1, a yeast homolog of the synaptic vesicle-associated membrane protein/synaptobrevin gene family: genetic interactions with the RAS and CAP genes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (10), 4338-4342 (1992)
MEDLINE 92262435
PUBMED 1316605
REFERENCE 10 (sites)
AUTHORS O'Connor,J.P. and Peebles,C.L.
TITLE PTA1, an essential gene of Saccharomyces cerevisiae affecting pre-tRNA processing
JOURNAL Mol. Cell. Biol. 12 (9), 3843-3856 (1992)
MEDLINE 92375052
PUBMED 1508188
REFERENCE 11 (sites)
AUTHORS De Virgilio,C., Burckert,N., Barth,G., Neuhaus,J.M., Boller,T. and Wiemken,A.
TITLE Cloning and disruption of a gene required for growth on acetate but not on ethanol: the acetyl-coenzyme A synthetase gene of Saccharomyces cerevisiae
JOURNAL Yeast 8 (12), 1043-1051 (1992)
MEDLINE 93190633
PUBMED 1363452
REFERENCE 12 (sites)
AUTHORS de Virgilio,C., Burckert,N., Neuhaus,J.M., Boller,T. and Wiemken,A.
TITLE CNE1, a Saccharomyces cerevisiae homologue of the genes encoding mammalian calnexin and calreticulin
JOURNAL Yeast 9 (2), 185-188 (1993)
MEDLINE 93220396
PUBMED 8465605
REFERENCE 13 (sites)
AUTHORS Teunissen,A.W., Holub,E., van der Hucht,J., van den Berg,J.A. and Steensma,H.Y.
TITLE Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerevisiae
JOURNAL Yeast 9 (4), 423-427 (1993)
MEDLINE 93289821
PUBMED 8511970
REFERENCE 14 (sites)
AUTHORS Ripmaster,T.L., Vaughn,G.P. and Woolford,J.L. Jr.
TITLE DRS1 to DRS7, novel genes required for ribosome assembly and function in Saccharomyces cerevisiae
JOURNAL Mol. Cell. Biol. 13 (12), 7901-7912 (1993)
MEDLINE 94067151
PUBMED 8247005
REFERENCE 15 (bases 99096 to 103271)
AUTHORS Keng,T., Clark,M.W., Storms,R.K., Fortin,N., Zhong,W., Ouellette,B.F., Barton,A.B., Kaback,D.B. and Bussey,H.
TITLE LTE1 of Saccharomyces cerevisiae is a 1435 codon open reading frame that has sequence similarities to guanine nucleotide releasing factors
JOURNAL Yeast 10 (7), 953-958 (1994)
MEDLINE 95076714
PUBMED 7985422
REFERENCE 16 (sites)
AUTHORS Sutrave,P., Shafer,B., Strathern,J. and Hughes,S.
TITLE Isolation, identification and characterisation of the FUN12 gene of Saccharomyces cerevisiae
JOURNAL Gene (1994) In press
REFERENCE 17 (bases 1 to 103682)
AUTHORS Bussey,H., Kaback,D.B., Zhong,W., Vo,D.T., Clark,M.W., Fortin,N., Hall,J., Ouellette,B.F.F., Keng,T., Barton,A.B., Su,Y., Davies,C.K. and Storms,R.K.
TITLE The nucleotide sequence of chromosome I from Saccharomyces cerevisiae
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92 (9), 3809-3813 (1995)
MEDLINE 95249563
PUBMED 7731988
REFERENCE 18 (sites)
AUTHORS Haarer,B. and Brown,S.
TITLE Identification of the yeast myosin gene that is similar to the yeast MYO2 gene
JOURNAL Unpublished (1992)
REFERENCE 19 (bases 1 to 103682)
AUTHORS Volckaert,G. and Valle,G.
JOURNAL Unpublished (1997)
REMARK genomic PCR verifications identified minor changes in the YAL063C and YAL062W intergenic region.
REFERENCE 20 (bases 1 to 103682)
AUTHORS Vo,D.
TITLE Direct Submission

```
  JOURNAL   Submitted (04-AUG-1994) Danh Vo, Biology, McGill University, 1205
            Docteur Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada
REFERENCE   21 (bases 1 to 103682)
  AUTHORS   Vo,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-1996) Danh Vo, Biology, McGill University, 1205
            Docteur Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada
  REMARK    Sequence update which affects many coding regions
REFERENCE   22 (bases 1 to 103682)
  AUTHORS   Jia,Y. and Cherry,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1997) Department of Genetics, Stanford
            University, Saccharomyces Genome Database, Stanford, CA 94305-5120,
            USA
COMMENT     On Feb 25, 1998 this sequence version replaced gi:1326053.

            Curated by:
            Saccharomyces Genome Database
            URL: http://genome-www.stanford.edu/
            e-mail: yeast-curator@genome.stanford.edu.
FEATURES             Location/Qualifiers
     source          1. .103682
                     /organism="Saccharomyces cerevisiae"
                     /strain="S288C/AB972"
                     /db_xref="taxon:4932"
                     /chromosome="I"
                     /map="1L; subtelomeric region to LTE1"
     gene            complement(4645. .6426)
                     /gene="SEO1"
     CDS             complement(4645. .6426)
                     /gene="SEO1"
                     /note="YAL067C; Suppressor of sulfoxyde ethionine
                     resistance; similar to S. cerevisiae Ycr28p Swiss-Prot
                     Accession Number P25621, with homology to allantoate
                     permease"
                     /citation=[8]
                     /codon_start=1
                     /product="Seo1p: putative membrane protein"
                     /protein_id="AAC04968.1"
                     /db_xref="GI:595520"
                     /translation="MYSIVKEIIVDPYKRLKWGFIPVKRQVEDLPDDLNSTEIVTISN

  Query Match           24.0%;  Score 274;  DB 8;  Length 103682;
  Best Local Similarity  53.9%;  Pred. No. 5.1e-66;
  Matches  612;  Conservative    0;  Mismatches  515;  Indels    9;  Gaps    2;

Qy       1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60
Db   32565 ATGAGAGCTTTGGCATATTTCAAGAAGGGTGATATTCACTTCACTAATGATATCCCTAGG 32624

Qy      61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120
Db   32625 CCAGAAATCCAAACCGACGATGAGGTTATTATCGACGTCTCTTGGTGTGGGATTTGTGGC 32684

Qy     121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180
Db   32685 TCGGATCTTCACGAGTACTTGGAT------GGTCCAATCTTCATGCCTAAAGATGGAGAG 32738

Qy     181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
Db   32739 TGCCATAAATTATCCAACGCTGCTTTACCTCTGGCAATGGGCCATGAGATGTCAGGAATT 32798

Qy     241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
Db   32799 GTTTCCAAGGTTGGTCCTAAAGTGACAAAGGTGAAGGTTGGCGACCACGTGGTCGTTGAT 32858

Qy     301 GCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGG 360
Db   32859 GCTGCCAGCAGTTGTGCGGACCTGCATTGCTGGCCACACTCCAAATTTTACAATTCCAAA 32918

Qy     361 CTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGT 420
Db   32919 CCATGTGATGCTTGTCAGAGGGGCAGTGAAAATCTATGTACCCACGCCGGTTTTGTAGGA 32978

Qy     421 TTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTC 480
Db   32979 CTAGGTGTGATCAGTGGTGGCTTTGCTGAACAAGTCGTAGTCTCTCAACATCACATTATC 33038

Qy     481 AAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCC 540
Db   33039 CCGGTTCCAAAGGAAATTCCTCTAGATGTGGCTGCTTTAGTTGAGCCTCTTTCTGTCACC 33098

Qy     541 TGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGA 600
Db   33099 TGGCATGCTGTTAAGATTTCTGGTTTCAAAAAAGGCAGTTCAGCCTTGGTTCTTGGTGCA 33158

Qy     601 GGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTG 660
Db   33159 GGTCCCATTGGGTTGTGTACCATTTTGGTACTTAAGGGAATGGGGGCTAGTAAAATTGTA 33218

Qy     661 TGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTC 720
Db   33219 GTGTCTGAAATTGCAGAGAGAAGAATAGAAATGGCCAAGAAACTGGGCGTTGAGGTGTTC 33278

Qy     721 GATCCTTCTA---CATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAAC 777
Db   33279 AATCCCTCCAAGCACGGTCATAAATCTATAGAGATACTACGTGGTTTGACCAAGAGCCAT 33338

Qy     778 GAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATT 837
Db   33339 GATGGGTTTGATTACAGTTATGATTGTTCTGGTATTCAAGTTACTTTCGAAACCTCTTTG 33398

Qy     838 GTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATT 897
Db   33399 AAGGCATTAACATTCAAGGGGACAGCCACCAACATTGCAGTTTGGGGTCCAAAACCTGTC 33458

Qy     898 GGATTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTAC 957
Db   33459 CCATTCCAACCAATGGATGTGACTCTCCAAGAGAAAGTTATGACTGGTTCGATCGGCTAT 33518

Qy     958 ACCGTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGAC 1017
Db   33519 GTTGTCGAAGCCTTCGAAGAAGTTGTTCGTGCCATCCACAACGGAGACATCGCCATGGAA 33578

Qy    1018 AAAGCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTT 1077
Db   33579 GATTGTAAGCAACTAATCACTGGTAAGCAAAGGATTGAGGACGGTTGGGAAAAGGGATTC 33638

Qy    1078 AAACAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACGA 1133
Db   33639 CAAGAGTTGATGGATCACAAGGAATCCAACGTTAAGATTCTATTGACGCCTAACAA 33694

RESULT 2
CNS06I04/c
LOCUS       CNS06I04                915 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T3 end of clone AS0AA016G03 of library AS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL399674
VERSION     AL399674.1  GI:12154481
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus.
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
```

```
   PUBMED   11152876
REFERENCE   2  (bases 1 to 915)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
  MEDLINE   20584715
   PUBMED   11152880
REFERENCE   3  (bases 1 to 915)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .915
                     /organism="Saccharomyces bayanus"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA016G03"
                     /clone_lib="AS0AA"
                     /note="end : T3"
     misc_feature    complement(<24. .>914)
                     /note="similar to Saccharomyces cerevisiae ORF YAL061w [
                     FUN50 ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      194 a    288 c    241 g    186 t      6 others
ORIGIN

  Query Match           14.3%;  Score 163;  DB 11;  Length 915;
  Best Local Similarity   52.3%;  Pred. No. 8.6e-35;
  Matches  478;  Conservative    2;  Mismatches  418;  Indels   16;  Gaps    5;

Qy      92 TCAAAGTCAGCTATTGTGGAATCTGTGGCACGGACTTGAAAGAATTCACATATTCTGGAG 151
           || |  ||   |  || || || || || |||||| || | || | ||      |||  |
Db     910 TCGATATCGCATGGTGCGGTATTTGCGGAACGGACCTGCATGAGTACA------CTGACG 857

Qy     152 GTCCTGTTTTTTTCCCTAAACAAGGCACCAAGGACAAGATTTCGGGATACGAACTTCCTC 211
           | ||  | || |||||  |  | ||   || | || ||||              |||| |
Db     856 GCCCCATCTTCTTCCCAGAGGATGGGCACACGAACGAGATCAGCCACANACCCATTCCGC 797

Qy     212 TCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACAAGTG 271
               || || ||  |    || ||||| |   | || ||  | || || ||
Db     796 AGGCAATGGGCCACGAGATGGCGGGCACGGTTGCGAAAGTGGGTCCCGGGGTRACGCAGT 737

Qy     272 TGAAACCTGGTGACAGAGTCGCAGTTGAAGCTACGTCGCATTGCTCCGACAGATCGCGCT 331
           | ||    || ||||  || |  |||||  | |||      |||   ||| | |  || |
Db     736 TCAAGGTGGGCGACAAGGTGGTTGTTGAGCCCACGGGTACGTGCAGGGACCGCTACCGTT 677

Qy     332 ACAAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGA 391
              | ||  |  |   |||   | |   |||   || ||| |. || ||       |
Db     676 GGCCCACCTCGCCCAACGTGGACAAGGAGTGGTGTGCCGCTTGCAAAAGGGGCCTCTACA 617

Qy     392 ACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGT 451
           ||  |||  |||  ||| || | || |||   || |     || || || ||||| |||
Db     616 ACATCTGCTCGTATCTGGGCCTTTGTGGTGCCGGCGTGCAGAGTGGAGGGTTTGCAGAGC 557

Qy     452 ACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTG 511
             || ||      |||   ||      |||| |||| || |   | ||    ||  |||
Db     556 GGGTGGTGATAAACGAGTCGCATTGCTACAAGGTGCCCGAGTTCGTGCCATTGGACGTTG 497

Qy     512 GAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGC 571
             ||  || |  |||| |  | ||   |||||||||  |          |   ||| |
Db     496 CTGCTTTGATCCAGCCGCTGGCCGTGTGCTGGCATGCGATCAGCGTGTGTGACTTCAAAG 437

Qy     572 CTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTC 631
           | ||    || || ||| |  |||| |  || || |||||||| | ||| ||||| || |
Db     436 CCGGGTCCACTGCTCTGATCATTGGTGCCGGCCCCATCGGCCTGGGCACGATTCTAGCAC 377

Qy     632 TGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGT 691
           |  | |    |  |    || | ||||     || ||||| ||   | | ||||| |||
Db     376 TCAATGCAGCTGGTTGCAGCGACATTGCCGTCTCAGAGCCTGCTAAGGTGAGAAGAGAGT 317

Qy     692 TTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACA------TGTGACGACGCAA 745
           | ||  || |  |||| ||     || ||||| ||  || ||        | | |||   |
Db     316 TGGCTGAGAAGATGGGTGCCAGGGTTTTCGACCCGACTGCACACGCCGCTAAGGACAGCA 257

Qy     746 ATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCA-TGCAGCCTTCGACTGC 804
           || | :| |   | |  | |  || |    ||| || | | | | || | | | ||||
Db     256 TTGATTWTTTGCGGTCGGTCGCTCCTGGCGGCGACGGGTACGACTACACCTTACAACTGA 197

Qy     805 TCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCC 864
           | | :  |       || || | | | ||  |          |     |   || | ||||
Db     196 TATARACTGGAAGTTACGTTGAACGCTTCCGTCCAATKTCTGACGTTTAGGGGCACCGCC 137

Qy     865 GTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCTGACTTAC 924
           || ||  ||||| | ||||| |||||    |   |||  ||||||    | ||
Db     136 GTGAACCTGGCCATGTGGGGTCACCACAAGGTGCAGTTCTCTCCAATGGACATCACGCTG 77

Qy     925 CAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGAAGTTGTC 984
           || |||| | |    || || ||||| ||||||||   | | |||||| |||    ||||
Db      76 CATGAGAGAAAGTACACGGGGTCCAT-TGCTACACACACCACGACTTCGAGG--CGTGTC 20

Qy     985 AAGGCCTTGGAAGA 998
            |||| ||||||||
Db      19 GAGGCTTTGGAAGA 6


RESULT 3
AX366887
LOCUS       AX366887               1041 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 11 from Patent WO0202766.
ACCESSION   AX366887
VERSION     AX366887.1  GI:18698175
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE   1
  AUTHORS   Bedzyk,L.A., Wang,T. and Ye,R.W.
  TITLE     Natural promoters for gene expression and metabolic monitoring in
            Bacillus species
  JOURNAL   Patent: WO 0202766-A 11 10-JAN-2002;
            E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Bacillus subtilis"
                     /db_xref="taxon:1423"
BASE COUNT      304 a    232 c    263 g    242 t
ORIGIN

  Query Match           7.3%;  Score 83;  DB 6;  Length 1041;
  Best Local Similarity   49.9%;  Pred. No. 4.1e-12;
  Matches  240;  Conservative    0;  Mismatches  235;  Indels    6;  Gaps    1;

Qy     407 TGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTG 466
           || | |||  ||| || |  || |  |||||||||  | || ||||||  |   | ||
Db     338 TGGGATTCCTCGGCTTAGCCGGCGGAGGCGGCGGTTTCTCTGAATACGTCTCTGTGGATG 397

Qy     467 AGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGC 526
           | || |   || |||| || || ||     |  | | ||    || || || ||||| |
Db     398 AAGAGCTTTTGTTCAAACTTCCTGATGAATTATCATATGAACAAGGCGCGCTCGTTGAAC 457

Qy     527 CTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCC 586
```

```
Db    458 CTTCTGCAGTTGCTCTATACGCTGTCCGCTCAAGCAAACTCAAAGCAGGCGACAAAGCGG 517

Qy    587 TGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATG 646

Db    518 CTGTATTCGGCTGCGGCCCGATCGGACTTCTTGTCATTGAAGCGCTGAAGGCTGCCGGTG 577

Qy    647 CGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGG 706

Db    578 CAACTGATATTTACGCTGTTGAGCTTTCTCCTGAACGCCAGCAAAAAGCTGAGGAGCTTG 637

Qy    707 GCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGG 766

Db    638 GCGCGATCATCGTTGATCCGTCTAAAACAGACGATGTAGTCGCTG------AGATTGCAG 691

Qy    767 TGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCA 826

Db    692 AACGTACAGGAGGCGGTGTTGACGTAGCATTCGAAGTCACTGGTGTCCCAGTGGTGTTAC 751

Qy    827 CCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAG 886

Db    752 GACAAGCCATCCAGTCCACTACAATTGCCGGTGAAACCGTCATCGTCAGCATTTGGGAAA 811

Qy    887 A 887

Db    812 A 812
```

```
RESULT 4
AB007638/c
LOCUS       AB007638               16585 bp    DNA     linear   BCT 13-FEB-1999
DEFINITION  Bacillus subtilis genomic DNA containing gutA to cotA region, 48
            degree.
ACCESSION   AB007638
VERSION     AB007638.1  GI:2522006
KEYWORDS    .
SOURCE      Bacillus subtilis (strain:Marburg 168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE   1  (sites)
  AUTHORS   Kasahara,Y., Nakai,S., Ogasawara,N., Yata,K. and Sadaie,Y.
  TITLE     Sequence analysis of the groESL-cotA region of the Bacillus
            subtilis genome, containing the restriction/modification system
            genes
  JOURNAL   DNA Res. (1997) In press
REFERENCE   2  (bases 1 to 16585)
  AUTHORS   Sadaie,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1997) Yoshito Sadaie, National Institute of
            Genetics, Radioisotope Center; Yata 1111, Mishima, Shizuoka 411,
            Japan (E-mail:ysadaie@lab.nig.ac.jp, Tel:81-0559-81-6870,
            Fax:81-0559-81-6870)
FEATURES             Location/Qualifiers
     source          1. .16585
                     /organism="Bacillus subtilis"
                     /strain="Marburg 168"
                     /db_xref="taxon:1423"
                     /map="48 degree"
     gene            1. .19
                     /gene="gutB"
     CDS             <1. .19
                     /gene="gutB"
                     /EC_number="1.1.1.14"
                     /note="L-idiotol 2-dehydrogenase"
                     /codon_start=2
                     /transl_table=11
                     /product="sorbitol dehydrogenase"
                     /protein_id="BAA22869.1"
                     /db_xref="GI:4432758"
                     /translation="VYPNR"
     gene            82. .1484
                     /gene="gutA"
     RBS             82. .88
                     /gene="gutA"
     CDS             93. .1484
                     /gene="gutA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sugar transport protein"
                     /protein_id="BAA22759.1"
                     /db_xref="GI:2522008"
                     /translation="MIGTEKVTALKAKHQSKSSTGLSWFERIGYGFGDMSYNIIFQFV
                     NAYLLFYYTDVGGIQPAVIATLFLVVRVLDAIFDPIMGLILDKTNTRWGKARPYLLWV
                     AFPFALFTFLCFTTPHFGETGNMVYAYVTYILLGMSFSMQTIPVNSLTGRMTNSVEER
                     TVLTTTRMILVYIGILLSISCATPLATAIGGEDQAFGFQMTALIYAAVSIVLNLFSFF
                     TVRERIQPKKRKKQGIKKTLSVLFKNKPLLMLISSFLAFAIGFNIKLSTMVYYFTYNV
                     NHKEFVFMGTVLFFGAALISNLFIPFFSEKWGRKQVMIITAALSLISYAGLHFTPYSS
                     IPLIFIWLFASGFFTTPLNTLAWGMVADCVDYAEWKTGIRADGVVISSMSFINKLGVA
                     LAGSFSAIYLGIAGYVANTDQTVASLNAIKNMNALIPGFFILLSIILIAFYPLTEKRY
                     KHIISELEQRSAR"
     gene            1563. .2541
                     /gene="fruC"
     RBS             1563. .1574
                     /gene="fruC"
     CDS             1579. .2541
                     /gene="fruC"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sugar transport protein"
                     /protein_id="BAA22760.1"
                     /db_xref="GI:2522009"
                     /translation="MRSHSVVCIGELLIDFFCTDVDVDLMEGRQFLKSAGGAPANVSA
                     AIAKLGGDAAFSGKVGKDPFGYFLKRTLDAVHVDTSMLVMDEKAPTTLAFVSLKQNGE
                     RDFVFNRGADALFTLEDIDQEKLNEAKILHFGSATALLSDPFCSAYLRLMSIAKDNGQ
                     FISFDPNYREDLWRGRVSEFVSVAKKAIAVSDFVKVSDEELEIISGVKDHEKGVAILH
                     EIGANIVAVTLGKSGTLLSNGKDREIIPSIPVTSIDSTGAGDAFVGAALYQLANTDQI
                     QSVDADFVKLREIVAFANKVGALVCTKIGAIDALPSLDEIGVSL"
     gene            2722. .3420
                     /gene="ydjF"
     RBS             2722. .2727
                     /gene="ydjF"
     CDS             2737. .3420
                     /gene="ydjF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="membrane protein"
                     /protein_id="BAA22761.1"
                     /db_xref="GI:2522010"
                     /translation="MSIIGRFKDIMSANINALLDKAENPEKMVDQYLRNMNSDLAKVK
                     AETAAVMAEEQRAKREYHECQADMEKMESYAMKALQAGNESDARKFLERKTSLESKLS
                     ELQAANQIAATNAAQMRKMHDKLVSDIGELEARKNMIKAKWAVAKTQERMNKLGASVS
                     STSQSMSAFGRMEDKVNKALDQANAMAELNSAPQDDMADLSAKYDTGGSSQVDDELAA
                     LKAKMMLDK"
     gene            join(3472. .3479,3486. .4511)
                     /gene="ydjG"
     RBS             3472. .3479
                     /gene="ydjG"
     CDS             3486. .4511
                     /gene="ydjG"
                     /note="function unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA22762.1"
                     /db_xref="GI:2522011"
                     /translation="MIISYKCPNCGSDMAFDSETGSLSCSSCGRQDNIESLPKENIAA
                     RFSDDEAKEYQCKNCGAVLITEAETTATTCSFCGGAAILADRLSGHLAPAKVIPFTIS
                     KQEAEQAFRKWCKKGLLTPRGFMSADRIKSITGMYIPFWMFDLNSEVQVRANCTRVHQ
                     YEEGDYICTETEHFEAFRDINLDYLKIPVDASEKMKDELMDKLEPYSYEELKDFQTAY
                     LAGYIAEKYNYTDEELFPRAKEKISSYIDSYIHSTFSGYTSVNVRDKHIHTKNVNSFY
                     VLLPVWMVSYDYERAEHIFAMNGQTGKVVGKPPISRGKVAAWFSGIAGGTFLALKLVS
                     LMMGGGF"
     gene            4497. .4501
                     /gene="ydjH"
```

```
RBS             4497. .4501
                /gene="ydjH"
gene            4511. .5275
                /gene="ydjH"
CDS             4511. .5275
                /gene="ydjH"
                /note="function unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA22763.1"
                /db_xref="GI:2522012"
                /translation="MRGFFGKAIFVVLAVFIMMPLLGIEAVRASELQQHVYDRAHLLS
                KAEIGKLESLSAKLGAKRDTDFIIITTKSTNGEDIADYTGDFYDRYGKGSTAILTIDM
                TNREVFIAGFKKAEQYLDNSRLNSIRNTISSDLSNENYFEAFETYIQLSYKDMGIKPG
                VNPDNIFFTWWFQLIAAIAVGGIAVSIMLYHAGGKVTVNGSTYMDQRTSDVIDQYDTY
                IRTTVTRERKPSDKDSGSDGGVTKGGTSYSGSRGSF"
gene            5296. .6277
                /gene="ydjI"
RBS             5296. .5301
                /gene="ydjI"
CDS             5306. .6277
                /gene="ydjI"
                /note="function unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA22764.1"
                /db_xref="GI:2522013"
                /translation="MSFFRNQLANVVEWEEFRDDMIFYKWNNREIKKGSRLIIRPGQD
                AVFLNNGRVEGVFQDDGDYDIESEIIPFLSTLKGFKFGFNSGMRAEVLFVNTKEFTVR
                WGTKQAINIPAAGMPGGMPIRANGTFNFKVQDYISLIDKIAGVKDQYFVEDIKTRIIS
                ILDQLLMKWITREGKDMFNLQANAFDIAKGIQEDLDMQLISDGMTVTGFQIMSFNYPQ
                EVQDMITKNASYGMVGDVNRYQQISMTDGMASGKMSGSGAASDMAGMMMGMNMANQMM
                NQMNQNQQAQSSGPQSTGSGSKPNFCPNCGTKTGEANFCPNCGQKLV"
terminator      6291. .6323
                /note="stem-loop"
gene            complement(6324. .7365)
                /gene="ydjJ"
CDS             complement(6324. .7349)
                /gene="ydjJ"
                /note="function unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA22765.1"
                /db_xref="GI:2522014"
                /translation="MNEGYIIAGLLLLTAGMIDFLWTTLWLESGAGPITRCLSAWLWK
                GCRKISGDHAKVLSMAGPLLLCLTLVIWISLFWSGWVLIYSSDPHSLMETQSKEPASW
                SDRIYFSGYVMFTLGNGDLAPNGGLWKLVTIIETAQGLLTITFSVTYLISVLSAVNQK
                RSFAQSVLSLGHDGTEIVHNAWNGKDFHDIDFLLVAASSELGKLTAQHNAFPILHFYH
                STQHQESSIIAVAVLDEALTIFKYGIPEQYQPNQLHIKEARSSIKNYLDTVHTAYIHP
                AEQAPPEPDISKLQQSGIPALSKQTFQIAVNSIKERRQLLLGIIQAGARKWPVQEQAI
                GNAYSPK"
RBS             complement(7358. .7365)
                /gene="ydjJ"
gene            7922. .9355
                /gene="ydjK"
RBS             7922. .7927
                /gene="ydjK"
CDS             7934. .9355
                /gene="ydjK"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="metabolite transport protein"
                /protein_id="BAA22766.1"
                /db_xref="GI:2522015"
                /translation="MNKQGNQMSFLRTIILVSTFGGLLFGYDTGVLNGALPYMGEPDQ
                LNLNAFTEGLVTSSLLFGAALGAVFGGRMSDFNGRRKNILFLAVIFFISTIGCTFAPN
                VTVMIISRFVLGIAVGGASVTVPAYLAEMSPVESRGRMVTQNELMIVSGQLLAFVFNA
                ILGTTMGDNSHVWRFMLVIASLPALFLFFGMIRMPESPRWLVSKGRKEDALRVLKKIR
                DEKRAAAELQEIEFAFKKEDQLEKATFKDLSVPWVRRIVFIGLGIAIVQQITGVNSIM
                YYGTEILRNSGFQTEAALIGNIANGVISVLATFVGIWLLGRVGRRPMLMTGLIGTTTA
                LLLIGIFSLVLEGSPALPYVVLSLTVTFLAFQQGAISPVTWLMLSEIFPLRLRGLGMG
                VTVFCLWMVNFAVSFTFPILLAAIGLSTTFFIFVGLGICSVLFVKRFLPETKGLSLEQ
                LEENFRAYDHSGAKKDSGAEVIG"
terminator      9360. .9393
                /note="stem-loop"
gene            complement(9403. .10457)
                /gene="ydjL"
CDS             complement(9403. .10443)
                /gene="ydjL"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="dehydrogenase"
```

Query Match 7.3%; Score 83; DB 1; Length 16585;
Best Local Similarity 49.9%; Pred. No. 5.6e-12;
Matches 240; Conservative 0; Mismatches 235; Indels 6; Gaps 1;

```
Qy     407 TGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTG 466
           || | |||  ||| || |  || | .  ||||||||||  | || ||||||  |   | ||
Db   10106 TGGGATTCCTCGGCTTAGCCGGCGGAGGCGGCGGTTTCTCTGAATACGTCTCTGTGGATG 10047

Qy     467 AGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGC 526
           | || |   || |||| || || ||     |  |  | ||    || || || ||||| |
Db   10046 AAGAGCTTTTGTTCAAACTTCCTGATGAATTATCATATGAACAAGGCGCGCTCGTTGAAC 9987

Qy     527 CTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCC 586
           ||  | | |||||     | |||||          | | || |  | ||  | |  ||
Db    9986 CTTCTGCAGTTGCTCTATACGCTGTCCGCTCAAGCAAACTCAAAGCAGGCGACAAAGCGG 9927

Qy     587 TGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATG 646
             | ||  | || || ||||| |||     ||||   || ||| | |       ||
Db    9926 CTGTATTCGGCTGCGGCCCGATCGGACTTCTTGTCATTGAAGCGCTGAAGGCTGCCGGTG 9867

Qy     647 CGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGG 706
           |     | |||     |   ||||  |       |    ||    ||  |||| || |
Db    9866 CAACTGATATTTACGCTGTTGAGCTTTCTCCTGAACGCCAGCAAAAAGCTGAGGAGCTTG 9807

Qy     707 GCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGG 766
           ||||     |  | ||||| |||| |   ||||| | |   ||||      ||  |   |
Db    9806 GCGCGATCATCGTTGATCCGTCTAAAACAGACGATGTAGTCGCTG------AGATTGCAG 9753

Qy     767 TGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCA 826
             |        |  ||  |  | | ||| |||||   | |||||| ||   |   ||
Db    9752 AACGTACAGGAGGCGGTGTTGACGTAGCATTCGAAGTCACTGGTGTCCCAGTGGTGTTAC 9693

Qy     827 CCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAG 886
             | ||     ||||   |  | | ||     ||||||  ||   | ||||||
Db    9692 GACAAGCCATCCAGTCCACTACAATTGCCGGTGAAACCGTCATCGTCAGCATTTGGGAAA 9633

Qy     887 A 887
           |
Db    9632 A 9632
```

```
RESULT 5
BSUB0004/c
LOCUS       BSUB0004              213190 bp    DNA     linear   BCT 26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 4 of 21): from 600701 to
            813890.
ACCESSION   Z99107 AL009126
VERSION     Z99107.1  GI:2632866
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE   1  (bases 1 to 213190)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
```

```
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
   PUBMED   9384377
REFERENCE   2  (bases 1 to 213190)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .213190
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            33. .740
                     /gene="ydfS"
     CDS             33. .740
                     /gene="ydfS"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12373.1"
                     /db_xref="GI:2632867"
                     /db_xref="SPTREMBL:P96697"
                     /translation="MIELEVVIRTVASFGLLLIAERILGKQTISQMTIFDFIAAITLG
                     AIAAGLAYNTSIKPHNMAISFSIFVLTIFLISFLSIKNRKLRKFFAGDPTVLIQNGKI
                     LESNMRKMRYTLDYLNQQLREKEIFNIEEVLFAILETNGQLTVLRKPQFRHVTKQDLM
                     IAVNQEQRLPIELIMDGEIIENNLKQNRLTESWLLEELRKRDIKVKETVYAVLLGNGD
                     IYVDQYKDHISVPMDKE"
     gene            complement(755. .1186)
                     /gene="ydfT"
     CDS             complement(755. .1186)
                     /gene="ydfT"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12374.1"
                     /db_xref="GI:2632868"
                     /db_xref="SPTREMBL:P96698"
                     /translation="MDFEKIRKWLEITNEYKQSDFWTNVLKYKAPEHFFDSEASTFVY
                     DFYQDEEYNFIIVEMPGVYEEELTIRLLSKTQLLIKGTITPVFPAEMEVLRERYYGEI
                     ERIIQLPEAAETHLLQIQLLNGLLHISYPRQVETVAFNKGL"
     gene            complement(1199. .1441)
                     /gene="ydgA"
     CDS             complement(1199. .1441)
                     /gene="ydgA"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12375.1"
                     /db_xref="GI:2632869"
                     /db_xref="SPTREMBL:P96699"
                     /translation="MPYQINIANIKINGVTQNGNIDVGPTVHNSHTANSKYFGANFSL
                     GDLSPTSSLLNTGNIDSDVSDQDQIGNPSAPISNQI"
     gene            complement(1455. .1727)
                     /gene="ydgB"
     CDS             complement(1455. .1727)
                     /gene="ydgB"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12376.1"
                     /db_xref="GI:2632870"
                     /db_xref="SPTREMBL:P96700"
                     /translation="MPSTVINLYYLKINSISGNGSITIGEAAYNSPTNNQKSQGTNSS
                     FGDTSPTESVMENFLNDPDVNDQTSIGNSDTSNVNAPPIAPPPILD"
     gene            2026. .2613
                     /gene="ydgC"
     CDS             2026. .2613
                     /gene="ydgC"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12377.1"
                     /db_xref="GI:2632871"
                     /db_xref="SPTREMBL:P96701"
                     /translation="MKSKGVESRKRLLKAAANEFSVRGFHDAKVSEIVKKAGFTQPSF
                     YLYFQSKEAIFAELITDFHSRVRKLTESLLLENGLNTEDVSKRVLLAVETVFQFLDED
                     KDLTKIGFFLNPEAKQMKKDLAMVLKENLEAEQRLGYFHSELDMETVAECLVGMIEHL
                     TESFLLTGIKDPASLAAEVVNLLIYGMLPKGNDVR"
     gene            join(2610. .2954,3074. .3091)
                     /gene="ydgD"
     CDS             2610. .2954
                     /gene="ydgD"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12378.1"
                     /db_xref="GI:2632872"
                     /db_xref="SPTREMBL:P96702"
                     /translation="MISIMMKVSLAVFMLAGGIIKVSRVPFQVEHWRHYQYPLWFLTV
                     TGILEIAGALAMTAGIWNRYAAIGAGVLFVVLMAGAIHAHMFRARQSVIMAIQAMICL
                     IVSIMIIMGSYT"
     terminator      3074. .3091
                     /gene="ydgD"
     gene            3117. .3590
                     /gene="ydgE"
     CDS             3117. .3590
                     /gene="ydgE"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12379.1"
                     /db_xref="GI:2632873"
                     /db_xref="SPTREMBL:P96703"
                     /translation="MAFTLEDMTEEEFEAFRGMSVQNFAKQNITSGTWTEKEAFEKSE
                     QAYENMIPNGRDSSNHYFWNITNDQGERMGWLWLYADPLHPQKEAFIYSFGLYEAFRG
                     KGLAQLALQTLDERARELGAERLALHVFAHNETAVYLYQKMGYAMTNIRMRKQLC"
     gene            complement(3705. .5393)
                     /gene="expZ"
     terminator      complement(3705. .3722)
                     /gene="expZ"
     CDS             complement(3750. .5393)
                     /gene="expZ"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="ATP-binding transport protein"
                     /protein_id="CAB12380.1"
                     /db_xref="GI:2632874"
                     /db_xref="SWISS-PROT:P39115"
                     /translation="MKEIVTLTNVSYEVKDQTVFKHVNASVQQGDIIGIIGKNGAGKS
                     TLLHLIHNDLAPAQGQILRKDIKLALVEQETAAYSFADQTPAEKKLLEKWHVPLRDFH
                     QLSGGEKLKARLAKGLSEDADLLLLDEPTNHLDEKSLQFLIQQLKHYNGTVILVSHDR
                     YFLDEAATKIWSLEDQTLIEFKGNYSGYMKFREKKRLTQQREYEKQQKMVERIEAQMN
                     GLASWSEKAHAQSTKKEGFKEYHRVKAKRTDAQIKSKQKRLEKELEKAKAEPVTPEYT
                     VRFSIDTTHKTGKRFLEVQNVTKAFGERTLFKNANFTIQHGEKVAIIGPNGSGKTTLL
                     NIILGQETAEGSVWVSPSANIGYLTQEVFDLPLEQTPEELFENETFKARGHVQNLMRH
                     LGFTAAQWTEPIKHMSMGERVKIKLMAYILEEKDVLILDEPTNHLDLPSREQLEETLS
                     QYSGTLLAVSHDRYFLEKTTNSKLVISNNGIEKQLNDVPSERNEREELRLKLETERQE
                     VLGKLSFMTPNDKGYKELDQAFNELTKRIKELDHQDKKD"
     gene            complement(5428. .7089)
                     /gene="ydgF"
     terminator      complement(5428. .5450)
                     /gene="ydgF"
     terminator      complement(5678. .5703)
                     /gene="ydgF"
     CDS             complement(5713. .7089)
                     /gene="ydgF"
                     /function="unknown"
                     /note="similar to amino acid ABC transporter (permease)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12381.1"
                     /db_xref="GI:2632875"
                     /db_xref="SWISS-PROT:P96704"
                     /translation="MTDDMTKDNINQQTLQRGLKNRHIQLIAIGGAIGTGLFLGSGKS
                     IHFAGPSILFAYMITGIICFLIMRSLGELLLSNLNYHSFVDFVQDYLGDMAAFITGWT
                     YWFCWISIAMADLTAVGLYTQYWLPGVPQWVPGLIALIILLIMNLATVKLFGELEFWF
                     ALIKVIAILALIVIGLVMIFKGFSTSSGVSSFTNLWSHGGLFPNGMHGFILSFQMVVF
                     AFVGIELVGLTAGETENPEKVIPKAINNIPVRVLLFYIGALLVIMSIYPWDIINPSES
                     PFVQVFVAVGIVGAASIINFVVLTSAASACNSAVFSTSRMVYSLAKDHNAPESMAKLT

  Query Match          7.3%;  Score 83;  DB 1;  Length 213190;
  Best Local Similarity  49.9%;  Pred. No. 7.5e-12;
  Matches  240;  Conservative    0;  Mismatches  235;  Indels    6;  Gaps    1;

Qy    407 TGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTG 466
Db  77631 TGGGATTCCTCGGCTTAGCCGGCGGAGGCGGCGGTTTCTCTGAATACGTCTCTGTGGATG 77572

Qy    467 AGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGC 526
Db  77571 AAGAGCTTTTGTTCAAACTTCCTGATGAATTATCATATGAACAAGGCGCGCTCGTTGAAC 77512

Qy    527 CTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCC 586
Db  77511 CTTCTGCAGTTGCTCTATACGCTGTCCGCTCAAGCAAACTCAAAGCAGGCGACAAAGCGG 77452

Qy    587 TGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATG 646
Db  77451 CTGTATTCGGCTGCGGCCCGATCGGACTTCTTGTCATTGAAGCGCTGAAGGCTGCCGGTG 77392

Qy    647 CGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGG 706
Db  77391 CAACTGATATTTACGCTGTTGAGCTTTCTCCTGAACGCCAGCAAAAAGCTGAGGAGCTTG 77332

Qy    707 GCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGG 766
Db  77331 GCGCGATCATCGTTGATCCGTCTAAAACAGACGATGTAGTCGCTG------AGATTGCAG 77278

Qy    767 TGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCA 826
Db  77277 AACGTACAGGAGGCGGTGTTGACGTAGCATTCGAAGTCACTGGTGTCCCAGTGGTGTTAC 77218

Qy    827 CCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAG 886
Db  77217 GACAAGCCATCCAGTCCACTACAATTGCCGGTGAAACCGTCATCGTCAGCATTTGGGAAA 77158

Qy    887 A 887
          |
Db  77157 A 77157


RESULT 6
AX312470
LOCUS       AX312470               339 bp    DNA     linear   PAT 14-DEC-2001
DEFINITION  Sequence 5455 from Patent WO0190366.
ACCESSION   AX312470
VERSION     AX312470.1  GI:17897465
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Leach,M.D. and Shimkets,R.A.
  TITLE     Human polynucleotides and polypeptides encoded thereby
  JOURNAL   Patent: WO 0190366-A 5455 29-NOV-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       78 a     70 c    101 g     90 t
ORIGIN

  Query Match          6.1%;  Score 69.8;  DB 6;  Length 339;
  Best Local Similarity  50.4%;  Pred. No. 2e-08;
  Matches  170;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy    254 GCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAGCTACGTCGCATT 313
Db      2 GCCCTGGTGTGAAAAACTTGAAAGTGGGAGACAAGGTAGTTGTCGAGCCCACAGGTACAT 61

Qy    314 GCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCT 373
Db     62 GCAGAGACCGGTATCGTTGGCCCCTGTCGCCAAACGTTGACAAGGAATGGTGCGCTGCTT 121

Qy    374 GCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCA 433
Db    122 GCAAAAAGGGCTACTATAACATTTGTTCATATTTGGGGCTTTGTGGTGCGGGTGTGCAGA 181

Qy    434 GCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACT 493
Db    182 GCGGTGGATTTGCAGAACGTGTTGTGATGAACGAATCTCACTGCTACAAAGTACCGGACT 241

Qy    494 CGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTG 553
Db    242 TCGTGCCCTTAGACGTTGCAGCTTTGATTCAACCGTTGGCTGTGTGCTGGCATGCAATTA 301

Qy    554 AACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGT 590
Db    302 GAGTCTGCGAGTTCAAAGCAGGCTCTACGGCTTTGAT 338


RESULT 7
CNS06K6L/c
LOCUS       CNS06K6L               717 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T7 end of clone AT0AA002E01 of library AT0AA from strain CBS 4311
            of Saccharomyces servazzii, sequence tagged site.
ACCESSION   AL402499
VERSION     AL402499.1  GI:12161549
KEYWORDS    STS.
SOURCE      Saccharomyces servazzii.
  ORGANISM  Saccharomyces servazzii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 717)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
```

```
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 717)
  AUTHORS   Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 7.
            Saccharomyces servazzii
  JOURNAL   FEBS Lett. 487 (1), 47-51 (2000)
  MEDLINE   20584717
   PUBMED   11152882
REFERENCE   3  (bases 1 to 717)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Saccharomyces servazzii"
                     /strain="CBS 4311"
                     /db_xref="taxon:27293"
                     /clone="AT0AA002E01"
                     /clone_lib="AT0AA"
                     /note="end : T7"
     misc_feature    complement(<10. .>333)
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]
                     similar to Saccharomyces cerevisiae ORF YAL061w [ FUN50 ;
                     similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      222 a    146 c    126 g    211 t     12 others
ORIGIN

  Query Match             5.9%;  Score 67.4;  DB 11;  Length 717;
  Best Local Similarity   53.1%;  Pred. No. 1e-07;
  Matches  171;  Conservative    0;  Mismatches  142;  Indels    9;  Gaps    1;

Qy       1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60
             ||||  |||||||  |||||       |   ||||| |  | | | || |   | || |||
Db     333 ATGAGAGGTTTAGCTTATTTTAAAGAAGGTGATATACATTTCACAGATAGTTTGCCAGAA 274

Qy      61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120
           ||  || | |     ||  |   | |    || |    |   | |   ||| | |||||
Db     273 CCAAAGCTTACTTCCCCTGATAGTCTACTTATAAGCCCCTCGTTTGTAGGATTGTGTGGT 214

Qy     121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180
           |  |||||| | ||           |  |   |||  | ||||||||||||| | |||  |
Db     213 AGTGACTTGCATGAG---------GCAAGCCATCCAATCTTTTTCCCTAAAGATGGCCAC 163

Qy     181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
           |   | ||| |    | | | |||||||| |    || |||||| ||  | || ||
Db     162 AGACATAAGCTAAGTGAAAATGAACTTCCCCARGCTCTTGGACACGAGGTRAGTGGTTAT 103

Qy     241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
           |||   |   ||||  | |  || ||     | |||   || ||| ||||||  || |||
Db     102 GTGAAAGCTATTGGACCAGCCGTGACTCAATTTAAAGTAGGGGACCGAGTCGTCGTCGAA 43

Qy     301 GCTACGTCGCATTGCTCCGACA 322
           || |   |   |||    || |
Db      42 GCAAGTGCAANTTGTGTTGATA 21


RESULT 8
AY133848
LOCUS       AY133848               1126 bp    mRNA    linear   PLN 07-AUG-2002
DEFINITION  Arabidopsis thaliana clone U10156 putative sorbitol dehydrogenase
            (At5g51970) mRNA, complete cds.
ACCESSION   AY133848
VERSION     AY133848.1  GI:22136875
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1126)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
            Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yu,G., Yuan,S., Bowser,L., Carninci,P., Chen,H., Cheuk,R.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
            Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Open Reading Frame (ORF) Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1126)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
            Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K.,
            Wong,C., Wu,H.C., Yu,G., Yuan,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
            Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2002) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     The RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
            Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W., Lee,J.M.,
            Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yu,G., Yuan,S., Bowser,L., Chen,H., Cheuk,R., Jones,T.,
            Karlin-Neumann,G., Kim,C., Lam,B., Lin,J., Miranda,M., Nguyen,M.,
            Palm,C.J., Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and
            Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
            to this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
            /PGEC) contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1126
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="U10156"
                     /note="This clone is in pUNI 51.
                     ecotype: Columbia"
     gene            1. .1126
                     /gene="At5g51970"
     CDS             1. .1095
                     /gene="At5g51970"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative sorbitol dehydrogenase"
```

```
                     /protein_id="AAM91782.1"
                     /db_xref="GI:22136876"
                     /translation="MGKGGMSQGEGSKVEEENMAAWLVGINTLKIQPFLLPSVGPHDV
                     RVRMKAVGICGSDVHYLKTMRCADFVVKEPMVIGHECAGIIEEVGEEVKHLVVGDRVA
                     LEPGISCWRCNLCREGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPENVSLEE
                     GAMCEPLSVGVHACRRAEVGPETNVLVMGAGPIGLVTMLAARAFSVPRIVIVDVDENR
                     LAVAKQLGADEIVQVTTNLEDVGSEVEQIQKAMGSNIDVTFDCAGFNKTMSTALAATR
                     CGGKVCLVGMGHGIMTVPLTPAAAREVDVVGVFRYKNTWPLCLEFLTSGKIDVKPLIT
                     HRFGFSQKEVEDAFETSARGSNAIKVMFNL"
     3'UTR           1096. .1126
                     /gene="At5g51970"
BASE COUNT      296 a    200 c    311 g    319 t
ORIGIN

  Query Match             5.9%;  Score 67;  DB 8;  Length 1126;
  Best Local Similarity   47.1%;  Pred. No. 1.4e-07;
  Matches  248;  Conservative    0;  Mismatches  270;  Indels    9;  Gaps    1;

Qy     357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
           | ||   || |    |||| |   |||       |||   ||| |     |||  |||| 
Db     327 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 386

Qy     417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
           |           |     |||  ||| ||  |  | || ||  ||  || ||||| 
Db     387 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 446

Qy     477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
           ||||  |||| ||    |        ||     |||||| ||  ||| ||  ||  ||| 
Db     447 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 506

Qy     537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
           ||     ||||||  |   || |||    |    ||||     |  |   ||||  | || 
Db     507 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 566

Qy     597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
           ||  || ||||| |||||||  || ||  | |||   |  |       || |   | ||| 
Db     567 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 626

Qy     657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
           |||   |   ||    |   || | |   |   | |||||  | || ||||| || | 
Db     627 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 686

Qy     717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
           | | |  |     |||    ||   |   ||| || |    |||  |     |    |||| 
Db     687 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 746

Qy     777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
           ||         || ||| | | ||||||||  | ||| ||    | ||  |  |
Db     747 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 806

Qy     828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
           ||  || | |  ||||  |   |   || |  | |     ||| |
Db     807 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 853


RESULT 9
AF370161
LOCUS       AF370161               1269 bp    mRNA    linear   PLN 23-APR-2002
DEFINITION  Arabidopsis thaliana putative sorbitol dehydrogenase (At5g51970)
            mRNA, complete cds.
ACCESSION   AF370161
VERSION     AF370161.1  GI:13877796
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1269)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1269)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,
            Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
            Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
            Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
            Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L.,
            Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E.,
            Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J.,
            Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1269
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL05-10-F02 (R10156)"
                     /note="This clone is in a modified pBluescript vector
                     (Lambda ZAP) as a XhoI/SstI insert.
                     ecotype: Columbia"
     gene            1. .1269
                     /gene="At5g51970"
     5'UTR           1. .52
                     /gene="At5g51970"
     CDS             53. .1147
                     /gene="At5g51970"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative sorbitol dehydrogenase"
                     /protein_id="AAK43976.1"
                     /db_xref="GI:13877797"
                     /translation="MGKGGMSQGEGSKVEEENMAAWLVGINTLKIQPFLLPSVGPHDV
                     RVRMKAVGICGSDVHYLKTMRCADFVVKEPMVIGHECAGIIEEVGEEVKHLVVGDRVA
                     LEPGISCWRCNLCREGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPENVSLEE
                     GAMCEPLSVGVHACRRAEVGPETNVLVMGAGPIGLVTMLAARAFSVPRIVIVDVDENR
                     LAVAKQLGADEIVQVTTNLEDVGSEVEQIQKAMGSNIDVTFDCAGFNKTMSTALAATR
                     CGGKVCLVGMGHGIMTVPLTPAAAREVDVVGVFRYKNTWPLCLEFLTSGKIDVKPLIT
                     HRFGFSQKEVEDAFETSARGSNAIKVMFNL"
     3'UTR           1148. .1269
                     /gene="At5g51970"
BASE COUNT      354 a    218 c    338 g    359 t
ORIGIN

  Query Match             5.9%;  Score 67;  DB 8;  Length 1269;
  Best Local Similarity   47.1%;  Pred. No. 1.4e-07;
  Matches  248;  Conservative    0;  Mismatches  270;  Indels    9;  Gaps    1;

Qy     357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
           | ||   || |    |||| |   |||       |||   ||| |     |||  |||| 
Db     379 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 438
```

```
Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
Db    439 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 498

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
Db    499 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 558

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
Db    559 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 618

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
Db    619 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 678

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
Db    679 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 738

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
Db    739 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 798

Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
Db    799 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 858

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
Db    859 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 905
```

RESULT 10
AY085213
LOCUS AY085213 1347 bp mRNA linear PLN 25-JUN-2002
DEFINITION Arabidopsis thaliana clone 13875 mRNA, complete sequence.
ACCESSION AY085213
VERSION AY085213.1 GI:21403923
KEYWORDS FLI_CDNA.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
TITLE Full-length messenger RNA sequences greatly improve genome annotation
JOURNAL Genome Biol. (2002) In press
REFERENCE 2 (bases 1 to 1347)
AUTHORS Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
TITLE Full-Length cDNA from Arabidopsis thaliana
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1347)
AUTHORS Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or LAer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.
FEATURES Location/Qualifiers
source 1. .1347
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="13875"
CDS 136. .1230
/codon_start=1
/product="sorbitol dehydrogenase-like protein"
/protein_id="AAM62446.1"
/db_xref="GI:21553353"
/translation="MGKGGMSQGEGSKVEEENMAAWLVGINTLKIQPFLLPSVGPHDV
RVRMKAVGICGSDVHYLKTMRCADFVVKEPMVIGHECAGIIEEVGEEVKHLVVGDRVA
LEPGISCWRCNLCREGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPENVSLEE
GAMCEPLSVGVHACRRAEVGPETNVLVMGAGPIGLVTMLAAQAFSVPRIVIVDVDENR
LAVAKQLGADEIVQVTTNLEDVGSEVEQIQKAMGSNIDVTFDCAGFNKTMSTALAATR
CGGKVCLVGMGHGIMTVPLTPAAAREVDVVGVFRYKNTWPLCLEFLTSGKIDVKPLIT
HRFGFSQKEVEDAFETSARGSNAIKVMFNL"
BASE COUNT 358 a 235 c 354 g 400 t
ORIGIN

Query Match 5.9%; Score 67; DB 8; Length 1347;
Best Local Similarity 47.1%; Pred. No. 1.4e-07;
Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

```
Qy    357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
Db    462 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTATGTCCAGAAATGAAGTTCTT 521

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
Db    522 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAGGTGGTTCATCCTGCGGATCTATG 581

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
Db    582 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 641

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
Db    642 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 701

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
Db    702 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCAGGCTTTCAGTGTGCCTAGAAT 761

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
Db    762 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 821

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
Db    822 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 881

Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
Db    882 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 941

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
Db    942 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 988
```

RESULT 11
AB025969
LOCUS AB025969 1457 bp mRNA linear PLN 04-APR-2000
DEFINITION Prunus persica mRNA for NAD-dependent sorbitol dehydrogenase, complete cds.
ACCESSION AB025969

```
VERSION     AB025969.1  GI:7416845
KEYWORDS    NAD-dependent sorbitol dehydrogenase.
SOURCE      Prunus persica fruit cDNA to mRNA.
  ORGANISM  Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 1457)
  AUTHORS   Niwa,N. and Yamada,K.
  TITLE     Prunus persica mRNA for NAD-dependent sorbitol dehydrogenase,
            complete cds
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 1457)
  AUTHORS   Niwa,N. and Yamada,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1999) Kunio Yamada, National Research Institute
            of Vegetables, Ornamental plants and Tea, Dep. of Floriculture,
            Lab. of Postharvest Technology; 360 Kusawa, Ano-cho, Age-gun, Mie
            514-2392, Japan (E-mail:yamakuni@nivot.affrc.go.jp,
            Tel:+81-59-268-4664)
FEATURES             Location/Qualifiers
     source          1. .1457
                     /organism="Prunus persica"
                     /db_xref="taxon:3760"
                     /tissue_type="fruit"
     CDS             197. .1300
                     /codon_start=1
                     /product="NAD-dependent sorbitol dehydrogenase"
                     /protein_id="BAA94084.1"
                     /db_xref="GI:7416846"
                     /translation="MGKGGMSSQGGALEAHVEQENMAAWLVGVNTLRIQPFKLPTVGP
                     NDVRVKIKAVGICGSDVHYLKTMKCADFVVQEPMVIGHECAGIVDEVGSLVKNLVPGD
                     RVALEPGISCWRCEQCKGGRYNLCPDMKFFATPPVHGSLANQIVHPADLCFKLPENVS
                     LEEGAMCEPLSVGVHACRRANIGPETNVLVIGAGPIGLVSVLSARAFGAARIVIVDVD
                     DERLSIAKSLGADDAVKVSTNPQDLEDEVSKISKAMKGGVDVSFDCVGFNKTMSTALS
                     ATRPGGKVCLVGMGHGVMTVPLTPAAAREVDVVGIFRYKNTWPLCLEFLRTGKIDVKP
                     LITHRFGFSQKEIEEAFETSARGGNAIKVMFNL"
BASE COUNT      364 a    333 c    365 g    395 t
ORIGIN
```

```
  Query Match             5.8%;  Score 65.8;  DB 8;  Length 1457;
  Best Local Similarity   46.9%;  Pred. No. 3.2e-07;
  Matches  248;  Conservative    0;  Mismatches  272;  Indels    9;  Gaps    1;

Qy     359 GGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCG 418
           ||   |||  |   ||| || ||||       ||   |||  |     |||  |||| ||
Db     534 GGAGATGTGAGCAGTGCAAGGGCGGGCGATACAATCTCTGCCCTGACATGAAGTTCTTCG 593

Qy     419 GTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGG 478
               | |  ||||   || ||  |  |  | ||  ||  ||  || | |
Db     594 CCACCCCACCAGTCCACGGCTCATTGGCAAATCAGATTGTGCACCCTGCAGATCTGTGCT 653

Qy     479 TCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTG 538
           |||||||  || ||     |       ||    || ||  ||   |||||| |   |||||
Db     654 TCAAGCTCCCGGAGAATGTGAGTTTGGAAGAAGGCGCCATGTGCGAGCCTCTGAGTGTTG 713

Qy     539 CCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAG 598
              |||||   |   || || |   ||   || |     |  | ||||||  | ||||
Db     714 GTGTTCATGCCTGTCGGCGAGCCAACATCGGCCCAGAAACAAATGTCCTGGTGATCGGAG 773

Qy     599 GAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTG 658
            ||| || ||||| ||||   |  | ||  ||   |  |     |   |||| || |||||
Db     774 CAGGGCCAATCGGGCTTGTGTCGGTGCTCTCTGCTCGTGCTTTTGGGGCGGCCAGAATTG 833

Qy     659 TGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAG--- 715
           |   |   ||   ||          |   |    ||||||||   || |||||  || |
Db     834 TCATTGTGGATGTGGATGATGAGCGCTTATCCATTGCAAAGTCTCTCGGCGCCGACGACG 893

Qy     716 ------TGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGC 769
                 ||    |     |    ||    | | |||   |||    |||    |     |
Db     894 CCGTCAAAGTCTCCACAAACCCTCAGGATTTAGAAGATGAAGTTTCTAAGATAAGTAAGG 953

Qy     770 CGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCA 829
```

```
           |   |||  |  |||  |  |||     |||  ||||||   |||  ||  |  |  ||  |  | |
Db     954 CCATGAAAGGTGGAGTAGATGTGAGCTTTGACTGCGTGGGTTTTAATAAAACTATGTCAA 1013

Qy     830 CCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGT 878
           |  |  |   ||||||  | ||     || |  |     |   ||||  | |
Db    1014 CGGCCCTCAGCGCCACCCGTCCGGGCGGCAAAGTTTGCCTCGTGGGCAT 1062
```

```
RESULT 12
AE008211
LOCUS       AE008211               10582 bp    DNA     linear   BCT 18-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 15 of
            187 of the complete sequence.
ACCESSION   AE008211 AE007870
VERSION     AE008211.1  GI:15158435
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon).
  ORGANISM  Agrobacterium tumefaciens str. C58 (Cereon)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 10582)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10582)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Approximately 800 bp of telomeric sequence missing from the left
            end of the chromosome and 200 bp missing from the right end.
FEATURES             Location/Qualifiers
     source          1. .10582
                     /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
                     /strain="C58"
                     /db_xref="taxon:181661"
     gene            complement(89. .661)
                     /gene="AGR_L_941GM"
     CDS             complement(89. .661)
                     /gene="AGR_L_941GM"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_941GMp"
                     /protein_id="AAK88702.1"
                     /db_xref="GI:15158436"
                     /translation="MGAKALTGVTYGGTNERTGFPPTRGEYDNLTRALSEAAGHAKTL
                     GLQFGIETVNRYENHLLNSAEQAVALVERIGADNIFIHLDTFHMNMEEKGIANGIIAA
                     RNHLKYMHMSESDRGTPGFGNVAWDEVFSALAAIRFNGVLALEGFAAMPVEMAGAIST
                     WRPVAANTEESLEKGLAFLRDKANQYRIFE"
     gene            complement(942. .1937)
                     /gene="AGR_L_265"
     CDS             complement(942. .1937)
                     /gene="AGR_L_265"
                     /note="(AE005211) putative permease component of transport
                     system, probably ribose specific"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_265p"
                     /protein_id="AAK88703.1"
                     /db_xref="GI:15158437"
                     /translation="MNSLIRKFNAIFGADMAGPVIALIAVMLIFGTFADNFLSLATFG
                     SVAFQLPELGLLTLAMLLPLLTGGINLSVTFAANLSGLAAAWVLQAHGGVDAPPSAFL
                     FACLAALVTGGAAGAMTGAAIAYTRAHPILVTLSMMIFLRGLGEFLTRGGDVSGFPAY
                     MAPLGHGTLLGLPIPLLIFIVCVGIWHVLLTRTKLGFGLLMIGSNIEAARYSGLNTRK
                     IQVLVYTLSGLMCAVAGIIMLARFNSVRVGHGESYLLITVLAAFLGGINPFGGFGRVL
                     PVFVALIVLQLLSSGLNLLGANQHLATALWGVLMIVVMAARGLFSSYFASLRKKV"
     gene            complement(1934. .2950)
                     /gene="AGR_L_266"
     CDS             complement(1934. .2950)
                     /gene="AGR_L_266"
```

/note="(AE005211) putative permease component of transport system, probably ribose specific"
/codon_start=1
/transl_table=11
/product="AGR_L_266p"
/protein_id="AAK88704.1"
/db_xref="GI:15158438"
/translation="MRRLILGHTTEFTLLVVMVLLCTGLSFATDRFLTISNAFDVLNV SAVNIIFAVGLLVVLISGGIDISFAVAASVVQYVTVLALNALGGGNWAEGFIIAGAVG LSLGFVNAFLVWRLNIISIVATISTFNIFFGLLMFFTKGVSIYDLPEWLSTRVVFYER EMPDGSWVELTLPVVVMILCCVATWFMISRTTVGRKLYAFGDNPEGARRFGINIGAMH YISFGWLGLMAGIAGLMQAHYAQEVVPNALYGRELDVLAATVLGGARLGGGKGSVIGC VLGVLMVSITQNGLNLMGVSPFAFKMIVGAIILVAITLSSTRFDKLLPSALRTSAKKG SAQR"

gene complement(2943. .4487)
/gene="AGR_L_268"

CDS complement(2943. .4487)
/gene="AGR_L_268"
/note="sugar ABC transporter, ATP-binding protein - Thermotoga maritima (strain MSB8)"
/codon_start=1
/transl_table=11
/product="AGR_L_268p"
/protein_id="AAK88705.1"
/db_xref="GI:15158439"
/translation="MPSATSAMLSRRKERYAIMSAEPLLSLKNVKVTFGGVRALKGVS FEVNPGEVHCLAGENGCGKSTVIKVITGVYKPESDAELYFDGKPIAAMTPTLAQSLGI QVIWQDLALFDEMTVAENIGFQYAVNGRFGLVDKRAIKNAAEKALARLGVSFDLDRPL KELPIAQRQIVAIARALVGEARLVFMDEPTASLTQSETDYLIDIVRNLSASGVAVVFV SHRLAEVLEISDRITVLRDGSLVGVFPVEGMTQSRVTELMTGRNFDSAVIAADHDDKP VVLSVRGLTRAGEFEDVSFDLRRGETLGITGLLGAGRTELALTLFGMHRPQSGEVLID GKKVDFHSNRDAIRVGVAYLSEDRLSLGLNQPQSIADNLVMASLDKLLSGGLISTSKK ADVVSRWISALGVKIGLPEDPIRTLSGGNQQRVAIAKWLAIGPKILILDAPTVGVDVG ARAGIFEIVRKLAAEGLSIIVISDEAPEVYFNTDRVIHMVEGRFHATYDPRRLSLTEL EAAIYA"

gene complement(4551. .5528)
/gene="AGR_L_271"

CDS complement(4551. .5528)
/gene="AGR_L_271"
/note="(AE005211) putative periplasmic binding protein, probable substrate ribose"
/codon_start=1
/transl_table=11
/product="AGR_L_271p"
/protein_id="AAK88706.1"
/db_xref="GI:15158440"
/translation="MKKIIAAAVGLSLALLSSAAFAEGPKVGVVVKIGGIPWFNAMEA GIKEQSNKLGVDGFMVGPTSADPALQVRAIEDLIAQKVDFIGVVPNDAKVLEPVLKKA QAAGIKVITHESPEQLGADWDFELASAKGFGEAHGKLLGEKMGGKGSYAVFVGSLTVP LHNAWADAAIAYIKANYPDMKLVGDRYGVAEDLDKSRSTALDLMSANADLKGFLAFGS QGPIGAGRAVEERRKDGKVFVIGPFSPGQGAKLIKSGALTGGFMWNPKQAGEVFITLA DRIAKGETPKAGDTIEGLGTIDPKGNTIVVDQLLKIDKESIDKLVSMGL"

gene 5768. .6847
/gene="AGR_L_273"

CDS 5768. .6847
/gene="AGR_L_273"
/note="fructose transport system repressor FruR PA3563 (imported) - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_273p"
/protein_id="AAK88707.1"
/db_xref="GI:15158441"
/translation="MTGISSKKATIYDLSILSGASASTVSAVLNGSWRKRRISEETAD KILSLAKAQRYTTNLQARGLRSSKSGLVGLLVPVYDNRFFSSMAQTFEGQARKRGLSP MVVSGRRDPEEERRTVETLIAYSIDALFIAGVTDPDGVHQVCARAALPHVNIDLPGKF ASSVISNNRHGAEILTAAILAHAAKGGSLGPDDVILFGGHDDHASRERIDGFHAAKAD YFGVEGGDDIEITGYSPHMTEMAFERFFGRRGRLPRCFFVNSSINFEGLLRFMGRHDG EAFGDIVVGCFDYDPFASFLPFPVYMIKPDIAQMLEKGFELLEENRTEPEVTIIEPQL IPPRTALEGPLDDIWDPVALRRMAK"

gene 6878. .7267
/gene="AGR_L_275"

CDS 6878. .7267
/gene="AGR_L_275"
/note="hypothetical protein PA0015 (imported) - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_275p"
/protein_id="AAK88708.1"
/db_xref="GI:15158442"
/translation="MTSTALDVVPICNRHHHCNGITMIEKLMKVLESGRDSPVLRFSL AKALMDTRQFEDAAQHLQEAIRQNPDYSAVWAALGECRERLGDHEGAAAAWIEGVAVA IRRGDIQAKRQMEVRFRRLQTRSLSAS"

gene 7391. .9274
/gene="AGR_L_278"

CDS 7391. .9274
/gene="AGR_L_278"
/note="ACETOIN CATABOLISM REGULATORY PROTEIN"
/codon_start=1
/transl_table=11
/product="AGR_L_278p"
/protein_id="AAK88709.1"
/db_xref="GI:15158443"
/translation="MGFSDHIKEIETVGQGRNTGRDAVVLESWRRCLDTYQLDPTKAR EAVIVSETRLREHRQQAEDLLHIARSGLERLYRQVAEQNYVLLLSDRQGVTVEFLGDP SFNNNLRKAGLYLGSEWSEPRAGTCAVGACIATGESLTIHQTDHFDVTHTPLSCTAAP IYDTQGALTAVLDISLLSSPILKTSQNMARHLVSSTVRRIELANLMASSRHDLVLRFA GAPEFLDVDPEAALSVDGSGRITGMTHAAARLLAASRGLDWRMPERLLGQRVEEFFEA GLDELASLTRSSLPQDRRIVVRDGTVLYAHAIEPQQRRSSVPLARSKPQPVTERIGGD GPAVRNLRRKVEKLSPARLPILLQGQTGTGKEHLARIIHDASGVSGRFVAVNCAAIPE QLIESELFGYLPGAFTGALAKGRKGLVEEAQGGTLFLDEIGDMPFAAQSRLLRVLAEG EVLPVGGSTPQKVDFRVISASHRPLNDMVRTGAFREDLYYRLNAAVLTLPALSERDDF PWLLDRILEKHGMPGRPLTLSNAARLTILHHRWPGNIRELDNAIAFAAALCDDGLITV EDLPEQLTPASSTPMIDSRGTDLRALLAACNGNISEAARRLGVDRTTLHRRMRRLGIG KPH"

gene 9311. .10504
/gene="AGR_L_281"

CDS 9311. .10504
/gene="AGR_L_281"
/note="(AE006323) 2,3-butanediol dehydrogenase (EC 1.1.1.4)"
/codon_start=1
/transl_table=11
/product="AGR_L_281p"
/protein_id="AAK88710.1"
/db_xref="GI:15158444"
/translation="MAGRKTGLPQNPAISGLRWRKPFWHGHCTFSGIKCHRGENMRAL RFHAAKDLRIEDIAEPKRPGPGQVLVRNRFVGICGTDLHEYSYGPIFIPTEPHPFTGA HGPQILGHEFGGVVEAIGDGVTSVNVGDRVSIQPLIMPRSGDYFADRGLFHLSTQLAL VGLSWDGGGMAEAALVNEYNVQKIPDEMTDEEAALVEPSAVAVYACDRGGVTAGNSVL VTGAGPIGMLTLLAARAAGATQLFVSDLNDARLELARNVIRDVITINPKRDNVGDVVR SATEGKVGCDVAIECVGNEHALKACVDAVRKQGVVVQTGLHPHENPIDWFQVTFRDLE IKGSWAYPTHYWPRVIRLIASGLLPATKIVTKRITLDTAVTEGFDALLDPAGTHLKIL IDLSK"

BASE COUNT 2210 a 3246 c 3137 g 1989 t
ORIGIN

Query Match 5.6%; Score 64.4; DB 1; Length 10582;
Best Local Similarity 47.5%; Pred. No. 1e-06;
Matches 225; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

```
Qy    396 CTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGT 455
          | |  ||   |||    |  |||  |  |  | | | |||||||  | ||  ||   ||
Db   9772 CAGCACGCAACTGGCACTGGTCGGCCTCAGCTGGGACGGCGGCGGCATGGCGGAAGCCGC 9831

Qy    456 CGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGC 515
            ||  |   ||  || || |    ||  | || |||  |||  |||||||    |  ||
Db   9832 ACTTGTCAACGAATACAACGTCCAGAAAATTCCGGACGAGATGACCGACGAAGAGGCCGC 9891

Qy    516 ACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGG 575
          ||||||  || ||     | ||||||    ||||   ||| |||| | |  |     | ||
Db   9892 ACTGGTGGAACCCTCGGCCGTTGCCGTTTATGCCTGTGATCGCGGTGGCGTTACCGCCGG 9951

Qy    576 TCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCA 635
          | | |  | |||||    || |  || || |||||| |    ||  |||| ||    |
```

```
Db   9952 TAATAGCGTTCTGGTCACGGGTGCCGGCCCGATCGGCATGTTGACGCTTCTGGCGGCCCG 10011

Qy    636 AGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGC 695

Db  10012 CGCCGCCGGCGCTACCCAGCTTTTTGTCTCCGATCTGAATGACGCCCGCCTCGAGCTTGC 10071

Qy    696 ---AAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGT 752

Db  10072 CCGCAATGTCATCCGTGATGTCATCACCATCAACCCCAAGCGCGATAATGTCGGTGATGT 10131

Qy    753 TCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGT 812

Db  10132 CGTCCGCTCGGCAACCGAAGGCAAGGTCGGCTGCGACGTGGCGATCGAATGCGTGGGTAA 10191

Qy    813 TCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGT 866

Db  10192 TGAACATGCGCTCAAGGCCTGCGTTGACGCCGTGCGCAAGCAGGGCGTCGTCGT 10245


RESULT 13
AE009402/c
LOCUS       AE009402               11742 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 172
            of 187 of the complete sequence.
ACCESSION   AE009402 AE008689
VERSION     AE009402.1  GI:17743239
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (U. Washington).
  ORGANISM  Agrobacterium tumefaciens str. C58 (U. Washington)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 11742)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     The genome of the natural genetic engineer Agrobacterium
            tumefaciens C58
  JOURNAL   Science 294 (5550), 2317-2323 (2001)
  MEDLINE   21608550
   PUBMED   11743193
REFERENCE   2  (bases 1 to 11742)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
            98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .11742
                     /organism="Agrobacterium tumefaciens str. C58 (U.
                     Washington)"
                     /strain="C58"
                     /db_xref="taxon:180835"
     gene            213. .3002
                     /gene="acnB"
                     /note="Atu4734"
     CDS             213. .3002
                     /gene="acnB"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="aconitate hydratase 2"
                     /protein_id="AAL45528.1"
                     /db_xref="GI:17743240"
                     /translation="MNLYLDYLAEIKSRETLGLAPKPIDDGALTAEIIELIEDTSSEY
                     RADALKFFIYNTLPGTTSAAGVKAAFLKQIILGEVLVPEITPSFALELLSHMKGGPSI
                     KVLLDVALGNDAAIDAKAGEVLKTQVFLYDADMFRLRDAYKEGSAIARDVLESYAKAE
                     FFTKLPDVDDEIKIVTFIAAEGDISTDLLSPGNQAHSRSDRQLHGQCMISPEAQAEIV
                     ALQQQHPDKRVMLIAEKGTMGVGSSRMSGVNNVALWTGKQASPYVPFVNYAPIVAGTN
                     GISPIFATTVDVTGGIGINLKNWVKKLGEDGKPILNNDGNPILEQKYSVETGTVLKID
                     AKNRKLRDENGTELVDLASSFTPQKMEFMKAGSSYAIVFGKKLQTFAAQTLGVEATQV
                     FAPNKEISIEGQGLTAVEKIFNRNAVGVTPGKVLHAGSDVRVKVNIVGSQDTTGLMTA
                     QELEAMAATVISPLVDGAYQSGCHTASVWDKKAQANTPKLMSFMHNFGVITGRDPKGV
                     YHSMTDVIHKVLNDITVDDRAIIIGGDSHTRMSKGVAFGADSGTVALALATGEATMPI
                     PQSVKVTFKGTMQPYMDFRDVVHATQAQMLQQHGDNVFQGRIIEVHIGTLLADQAFTF
                     TDWTAEMKAKASICISEDETLIESLEIAKSRIQIMIDKGMDNAAQTLKGLIDKADQRI
                     AEIRSGETPALKPDANAKYFAEVVVDLDIINEPMIADPDVNNNDVSRRYTHDTIRPVS
                     YYGATKKVDLGFVGSCMVHKGDVKIVAQMLRNMEKTEGKVEFKAPLVVAAPTYNIIDE
                     LKAEGDWEILQKYSGFEFDDVNPKTSNRTEYDNILYLERPGCNLCMGNQEKAAKGDTV
                     LATSTRLFQGRVVEDSAEKKGESLLASTPVVVLSAILGRTPSIEEYRSAVEGIDLTKF
                     APPKTTPIDSQSVHY"
     gene            complement(2999. .3196)
                     /gene="Atu4735"
     CDS             complement(2999. .3196)
                     /gene="Atu4735"
                     /note="putative; ORF located using Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL45529.1"
                     /db_xref="GI:17743241"
                     /translation="MKCYPTYASADNSARESPFNLLHQIIRNHLHTNENAAEIDFRCV
                     HSSDAGQLGRIHRPKPAKPDA"
     gene            3399. .5354
                     /gene="Atu4736"
     CDS             3399. .5354
                     /gene="Atu4736"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl-accepting chemotaxis protein"
                     /protein_id="AAL45530.1"
                     /db_xref="GI:17743242"
                     /translation="MRPFMSFLKNTKIKTKVVFVISLMSLMSLSGISYVSLQYKKTDA
                     VYSDFIGHEALAAVLNARTSGNLNALGMQILRASLNDPASSDFEAAVKTFRADRKQLE
                     ERQNKIMELVPARTDAARDILKGVVEVEEIGNQVITLMQAGKQAEAQQMALNVLRKIT
                     EVSPKISAGNEQLIQVMNDGRERLTASTNTTIWTGLITLALVSIALIALGLLVSSRGI
                     TTPIARLRARMESLAAGDTASEIDGMDRRDEVGQMAATVQAFRENAIARVRLENETEA
                     NRSMSEKDRIEREKQKAQEAADVQFAVNNLATALSRIAEGDVTYRIAQPFVPNLDGIR
                     GDFNRAAEQLQSTLTQVAQNARGIDAGANEIRSAADDLARRTEQQAAAVEETAAALEE
                     ITTTVKDSTKRAQEAGHLVGRAKIGAEKSGEVVQKAVSAMEQIATSANEISNIIGVID
                     EIAFQTNLLALNAGVEAARAGDAGKGFAVVAQEVRELAQRSANAAKEIKNLITTSNGQ
                     VQQGVQLVGETGRALELIVTEVQEINRHVAAISESAQEQSSGLQQINTAVNQMDQDTQ
                     KNAAMVEESTAASHGLAREASSLNRLIAQFKLSEAGYAETAAPMRTASAADRPAVSPA
                     RALGGRIRAAFSGNAALNTAADNWEEF"
     gene            5553. .5735
                     /gene="Atu4737"
     CDS             5553. .5735
                     /gene="Atu4737"
                     /note="putative; ORF located using Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL45531.1"
                     /db_xref="GI:17743243"
                     /translation="MIAAYIDEFIMFCAGLWMTLVGFGYLALSKNPAYQPRVVRHFKW
                     MGPMLIGIAIILAIAS"
     gene            5940. .6569
                     /gene="cat"
```

/note="Atu4738"
CDS 5940. .6569
/gene="cat"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="AAL45532.1"
/db_xref="GI:17743244"
/translation="MENYFESPFRGITLDKQVKSPNLVVGKYSYYSGYYHGHSFEDCA
RYLLPDEGADRLVIGSFCSIGSGAAFIMAGNQGHRNEWISTFPFFFMPEVPEFENAAN
GYLPAGDTVIGNDVWIGSEAIIMPGITVGDGAVIGTRALVTKDVEPYAIVGGNPAKTI
RKRFDDDSIALLLEMKWWGWPAERLKAAMPLMTSGNVAALYRFWRSDSL"
gene complement(6623. .8032)
/gene="Atu4739"
CDS complement(6623. .8032)
/gene="Atu4739"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="flavin-containing monooxygenase"
/protein_id="AAL45533.1"
/db_xref="GI:17743245"
/translation="MRRPMGAEHGHDRSRRTWKEKREAEAAELGYAEQPYVVIIGGGQ
GGIALGARLRQLGVPTIIIEKNERPGDSWRKRYKSLCLHDPVWYDHLPYIPFPENWPV
FTPKDKVGDWLEMYTRVMELNYWGSTTCKSAQYDETTGEWTVVVDRAGKEVVLRPKQL
VLATGMSGKANVPKFPGQDVFKGEQQHSSQHPGPDAYAGKKVVVIGSNNSAHDICAAL
WEAGADVTMLQRSSTHIVKSDSLMEIGLGDLYSERAVANGMTTRKADLIFASLPYRIM
HEFQVPIYDRIREQDAEFYAALEKAGFMLDFGDDESGLFMKYLRRGSGYYIDVGACDL
VIDGSIKLKSGVDVSHLTENAVVLKDGTELPADLVVYATGYGSMNGWAADLISRDVAD
KVGKCWGLGSDTTKDPGPWEGEQRNMWKPTRQDALWFHGGNLHQSRHYSQYLSLQLKA
RQAGLKTPVYGLQKPHHLS"
gene complement(8538. .9611)
/gene="Atu4740"
CDS complement(8538. .9611)
/gene="Atu4740"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="zinc-binding dehydrogenase"
/protein_id="AAL45534.1"
/db_xref="GI:17743246"
/translation="MRALRFHAAKDLRIEDIAEPKRPGPGQVLVRNRFVGICGTDLHE
YSYGPIFIPTEPHPFTGAHGPQILGHEFGGVVEAIGDGVTSVNVGDRVSIQPLIMPRS
GDYFADRGLFHLSTQLALVGLSWDGGGMAEAALVNEYNVQKIPDEMTDEEAALVEPSA
VAVYACDRGGVTAGNSVLVTGAGPIGMLTLLAARAAGATQLFVSDLNDARLELARNVI
RDVITINPKRDNVGDVVRSATEGKVGCDVAIECVGNEHALKACVDAVRKQGVVVQTGL
HPHENPIDWFQVTFRDLEIKGSWAYPTHYWPRVIRLIASGLLPATKIVTKRITLDTAV
TEGFDALLDPAGTHLKILIDLSK"
gene complement(9768. .11651)
/gene="acoR"
/note="Atu4741"
CDS complement(9768. .11651)
/gene="acoR"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAL45535.1"
/db_xref="GI:17743247"
/translation="MGFSDHIKEIETVGQGRNTGRDAVVLESWRRCLDTYQLDPTKAR
EAVIVSETRLREHRQQAEDLLHIARSGLERLYRQVAEQNYVLLLSDRQGVTVEFLGDP
SFNNNLRKAGLYLGSEWSEPRAGTCAVGACIATGESLTIHQTDHFDVTHTPLSCTAAP
IYDTQGALTAVLDISLLSSPILKTSQNMARHLVSSTVRRIELANLMASSRHDLVLRFA
GAPEFLDVDPEAALSVDGSGRITGMTHAAARLLAASRGLDWRMPERLLGQRVEEFFEA
GLDELASLTRSSLPQDRRIVVRDGTVLYAHAIEPQQRRSSVPLARSKPQPVTERIGGD
GPAVRNLRRKVEKLSPARLPILLQGQTGTGKEHLARIIHDASGVSGRFVAVNCAAIPE
QLIESELFGYLPGAFTGALAKGRKGLVEEAQGGTLFLDEIGDMPFAAQSRLLRVLAEG
EVLPVGGSTPQKVDFRVISASHRPLNDMVRTGAFREDLYYRLNAAVLTLPALSERDDF
PWLLDRILEKHGMPGRPLTLSNAARLTILHHRWPGNIRELDNAIAFAAALCDDGLITV
EDLPEQLTPASSTPMIDSRGTDLRALLAACNGNISEAARRLGVDRTTLHRRMRRLGIG

Query Match 5.6%; Score 64.4; DB 1; Length 11742;
Best Local Similarity 47.5%; Pred. No. 1e-06;
Matches 225; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

```
Qy   396 CTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGT 455
Db  9270 CAGCACGCAACTGGCACTGGTCGGCCTCAGCTGGGACGGCGGCGGCATGGCGGAAGCCGC 9211

Qy   456 CGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGC 515
Db  9210 ACTTGTCAACGAATACAACGTCCAGAAAATTCCGGACGAGATGACCGACGAAGAGGCCGC 9151

Qy   516 ACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGG 575
Db  9150 ACTGGTGGAACCCTCGGCCGTTGCCGTTTATGCCTGTGATCGCGGTGGCGTTACCGCCGG 9091

Qy   576 TCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCA 635
Db  9090 TAATAGCGTTCTGGTCACGGGTGCCGGCCCGATCGGCATGTTGACGCTTCTGGCGGCCCG 9031

Qy   636 AGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGC 695
Db  9030 CGCCGCCGGCGCTACCCAGCTTTTTGTCTCCGATCTGAATGACGCCCGCCTCGAGCTTGC 8971

Qy   696 ---AAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGT 752
Db  8970 CCGCAATGTCATCCGTGATGTCATCACCATCAACCCCAAGCGCGATAATGTCGGTGATGT 8911

Qy   753 TCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGT 812
Db  8910 CGTCCGCTCGGCAACCGAAGGCAAGGTCGGCTGCGACGTGGCGATCGAATGCGTGGGTAA 8851

Qy   813 TCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGT 866
Db  8850 TGAACATGCGCTCAAGGCCTGCGTTGACGCCGTGCGCAAGCAGGGCGTCGTCGT 8797
```

RESULT 14
AB015478
LOCUS AB015478 81347 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MSG15.
ACCESSION AB015478 BA000015
VERSION AB015478.1 GI:3241926
KEYWORDS .
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MSG15.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Kotani,H., Miyajima,N. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. VII. Sequence features of the regions of 1,013,767 bp covered by sixteen physically assigned P1 and TAC clones
JOURNAL DNA Res. 5 (5), 297-308 (1998)
MEDLINE 99087489
REFERENCE 2 (bases 1 to 81347)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MSG15
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant

protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
The 5' clone is MJM18 and the 3' clone is F17P19.

FEATURES Location/Qualifiers

source 1. .81347
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MSG15"
/clone_lib="Mitsui P1"

exon complement(1. .402)
/note="CDS is reported in Acc# AB025623
gene_id:MJM18.7
pir||T05129
strong similarity to unknown protein"
/number=1
/evidence=not_experimental

CDS join(2611. .3415,4019. .4413,4496. .5056)
/note="gene_id:MSG15.1"
/codon_start=1
/evidence=not_experimental
/product="mandelonitrile lyase-like protein"
/protein_id="BAB11041.1"
/db_xref="GI:10177728"
/translation="MTIISCITIHIFRNILFAILIFYGSCYCDKGTILISYKYNLILV
YISIFFFFLVKLETTFMKDATLAPKNASFDYIIIGGGTAGCALAATLSQNASVLVLER
GGSPYENPTATDMGNSVNTLLNNTPNSWSQLFISEDGVYNTRPRVLGGGSVINGGFYS
RAGNDYVEEAEWEMEEVEAAYEWVEKKLVFEPQVIEWQKAFKDGLLEAGESPDNGFTY
DHIYGTKIGGTIFDRAGHRHTAANLLEYANPNRIVVYLHASVHKVLFTTEEKPKAYEV
LFEDANGVFHKANLANKATNEVILSAGALGSPQLLMLSGVGPAVHLEAHGVNPLVLDQ
PMVGQGMADNPMNFVAIPSPQPVELSLIQAVGITKFDSYIEGLSGLSLSFDITRRFFD
GVLNLLNETSHTTSRKILTQSIAVLLKSFDVKLEVRMNGGLIFQKVDGPASKGHMKLR
NTNPRDNPSVTFNYYQEPEDLNKCVKGLNTIIRMINSKAFSKYKYPGVTARELLNLML
ALPINLRPRHVTSAFNLKQFCIDTVTSVWHYHGGCQVGKVVDKNYKVLGIDGLRVIDG
STFLKSPGTNPQATVMMLGR"

CDS join(6167. .6204,6527. .6560,6653. .6825,6914. .6965,
7061. .7128,7215. .7284)
/note="gene_id:MSG15.2"
/codon_start=1
/evidence=not_experimental
/product="DNA-directed RNA polymerase II subunit-like protein"
/protein_id="BAB11042.1"
/db_xref="GI:10177729"
/translation="MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPI
EAEDKVETEPVQRPRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAM
KELRQRKIPFTIRRYLPDGSFEEWGVDELIVEDSWKRQVGGD"

CDS complement(join(7728. .8294,8400. .8800,9212. .9853,
9966. .10047))
/note="gene_id:MSG15.3"
/codon_start=1
/evidence=not_experimental
/product="mandelonitrile lyase-like protein"
/protein_id="BAB11043.1"
/db_xref="GI:10177730"
/translation="MGLQILWLFRKIFIITYIFHASCSPDKAGYYSFMKDATLAPMFA
RFDYIIIGGGTSGCALAATLSQNASVLVLERGGAPYDNPTATDIENFATTLSNTSPKS
WSQLFISEDGVYNTRARVLGGGSVLNAGFYTRAGDEYVKETEWKTDEVEAAYEWVEKK
VAFQPPVLGWQTAFKDGLLEAGEFPYNGFTYDHIYGTKIGGTIFDRAGHRHTAADLLE
YANPGNIVVYLHASVHKILFTTKGRPRPKAYGVIFQDANGVLHKAELEKNSMNEVILS
AGAIGSPQLLMLSGIGPAAHLAAHGIKPLVLDHPMVGQGMGDNPMNAIFIPSPTPVEV
SLIQVVGITKFESYIEGASGVIFSYSWTRRFFDGVLNYLNESRTTSTTSPTLSTQSIT
DFFNPINPLLNATTRAGLILQKIAGPISRGHLELRNTNPDDNPSVRFNYYQEPEDLQI
CVEGINTIIKVINSKAFSKFKYPDATIHGLLDLMLSVPTNLRPRHITSMFNLRQFCID
TVMTIWHYHGGCQVGRVVDKNYRVLGIDSLRVIDGSTFLKSPGTNPQATVMMLGR"

CDS join(12204. .12401,12491. .12604)
/note="gene_id:MSG15.4
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11044.1"
/db_xref="GI:10177731"
/translation="MSGEVVLAARVYRDLLKAVVKHVGKEDHKSHFLDFVKQEFRKNA
NSEKINLARNYTYLLNSIHSHKDLLFSYNIAVDRTEEMKRVLNKSAASVGLRLPEVYE
P"

CDS join(13308. .13425,13521. .13588,13687. .14214,14299. .14486,
14580. .14772)
/note="gene_id:MSG15.5"
/codon_start=1
/evidence=not_experimental
/product="sorbitol dehydrogenase-like protein"
/protein_id="BAB11045.1"
/db_xref="GI:10177732"
/translation="MGKGGMSQGEGSKVEEENMAAWLVGINTLKIQPFLLPSVGPHDV
RVRMKAVGICGSDVHYLKTMRCADFVVKEPMVIGHECAGIIEEVGEEVKHLVVGDRVA
LEPGISCWRCNLCREGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPENVSLEE
GAMCEPLSVGVHACRRAEVGPETNVLVMGAGPIGLVTMLAARAFSVPRIVIVDVDENR
LAVAKQLGADEIVQVTTNLEDVGSEVEQIQKAMGSNIDVTFDCAGFNKTMSTALAATR
CGGKVCLVGMGHGIMTVPLTPAAAREVDVVGVFRYKNTWPLCLEFLTSGKIDVKPLIT
HRFGFSQKEVEDAFETSARGSNAIKVMFNL"

CDS complement(join(15138. .15282,15363. .15460,15557. .15616,
15695. .15814,15916. .16062,16155. .16256,16376. .16465,
16839. .17390))
/note="contains similarity to myosin heavy chain kinase
gene_id:MSG15.6"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11046.1"
/db_xref="GI:10177733"
/translation="MDFDLNGGNKRVFNRLGGGGGSTRPMAPTDTRQKVCFHWRAGRC
NRSPCPYLHRELPGPGPGQGQGPGYTNKRVAEESGFAGPSHRRGPGFNGNSSSSWGRF
GGNRTVTKTEKVCNFWVDGNCTYGDKCRYLHCWSKGESFALLTQLDGHEKLVSGIALP
SGSDKLYTGSKDETLRVWDCASGQCTGVLKLGGEIGCVLSEGPWLLVGMPNLVKAWNI
ETNADQSLSGPVGQVYSLVVGTDLLFAGTQDGSILAWRYNAATNCFEPSASLTGHTLA
VVTLYVGANRLYSGSMDKTIKVWSLDNLQCIQTLTDHSSVVMSLICWDQFLLSCSLDN
TVKIWAAIEGGNLEVTYTHKEEHGVLALCGVHDAEAKPVLLCACNDNTLRLYDLPSFT
ERGKIFAKQEIRAIQIGPGGIFFTGDGTGQVKVWKWCTEPTAALP"

CDS complement(18601. .19275)
/note="gene_id:MSG15.7"
/codon_start=1
/evidence=not_experimental
/product="AP2 domain transcription factor-like protein"
/protein_id="BAB11047.1"
/db_xref="GI:10177734"
/translation="MNPFYSTFPDSFLSISDHRSPVSDSSECSPKLASSCPKKRAGRK
KFRETRHPIYRGVRQRNSGKWVCEVREPNKKSRIWLGTFPTVEMAARAHDVAALALRG
RSACLNFADSAWRLRIPETTCPKEIQKAASEAAMAFQNETTTEGSKTAAEAEEEAAGEG
VREGERRAEEQNGGVFYMDDEALLGMPNFFENMAEGMLLPPPEVGWNHNDFDGVGDVS
LWSFDE"

CDS complement(join(20512. .20836,20925. .21052,21150. .21278,
21377. .21455,21539. .21702,21778. .21954,22033. .22155,
22230. .22312,22406. .22505,22590. .22607))
/note="gene_id:MSG15.8"
/codon_start=1
/evidence=not_experimental
/product="importin alpha subunit"
/protein_id="BAB11048.1"
/db_xref="GI:10177735"
/translation="MAWKTEVNEVSDDIIDGLWSDDPPLQLESVTKIRRITSQRDISC
VIRSGVVPRLVQLLKNQVFPKLQYEVAWALTNIAVDNPGVVVNNNAVPVLIQLIASPK
DYVREQAIWTLSNVAGHSIHYRDFVLNSGVLMPLLRLLYKDTTLRIATWALRNLCRGK
PHPAFDQVKPALPALEILLHSHDEDVLKNACMALCHLSEGSEDGIQSVIEAGFVPKLV
QILQLPSPVVLVPALLTIGAMTAGNHQQTQCVINSGALPIISNMLTRNHENKIKKCAC

```
                     WVISNITAGTKEQIQSVIDANLIPILVNLAQDTDFYMKKEAVWAISNMALNGSHDQIK
                     YMAEQSCIKQLCDILVYSDERTTILKCLDGLENMLKAGEAEKNSEDVNPYCLLIEDAE
                     GLEKISKLQMNKNDDIYEKAYKILVTNWFEEDDENNNNNVRCDDVDFQV"
     CDS             complement(23114. .24304)
                     /note="dbj|BAA78737.1
                     gene_id:MSG15.9
                     strong similarity to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB11049.1"
                     /db_xref="GI:10177736"
                     /translation="MSLFCNSIETLYPKPFLSHKSHFFTFSNSKFQNPKFTSFPLRHQ
                     WRIESCLKSASSSEIDMVRNKEGIFAPKEKKVVVLWDLDNKPPRGPPYEAATALRKVA
                     EKLGRVVEISAYANRHAFIHLPHWVVEERRERRNLDFMERKGEVTPIDPYICGVCGRK
                     CKTNLDLKKHFKQLHERERQKKVNRMRSLKGKKRQKFKERYVSGNEKYNEAARSLLTP
                     KVGYGLEAELRRAGVYVKTVEDKPQAADWAVKRQIQHSMTRGIDWLVLVSDDKDFSDM
                     LRKAREADLGTLVVSDMDRALGRHADLWVPWSGVEKGEIGEKDLIPGKRPRFEEDEVG
                     FGGNDELFSLSYDEDEKAEMTVESDGFGRFNVSAFSEDEWVEDEGDFVLSDSSDDDSD
                     DELF"
     CDS             complement(25680. .26378)
                     /note="contains similarity to AP2 domain transcription
                     factor
                     gene_id:MSG15.10"

  Query Match             5.6%;  Score 64.2;  DB 8;  Length 81347;
  Best Local Similarity   48.5%;  Pred. No. 1.4e-06;
  Matches  177;  Conservative    0;  Mismatches  188;  Indels    0;  Gaps    0;

Qy     357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
Db   13827 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 13886

Qy     417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
Db   13887 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 13946

Qy     477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
Db   13947 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 14006

Qy     537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
Db   14007 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 14066

Qy     597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
Db   14067 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 14126

Qy     657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
Db   14127 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 14186

Qy     717 GTTCG 721
Db   14187 GATTG 14191


RESULT 15
AE004831
LOCUS       AE004831               15088 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 392 of 529 of the complete
            genome.
ACCESSION   AE004831 AE004091
VERSION     AE004831.1  GI:9950347
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 15088)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
   PUBMED   10984043
REFERENCE   2  (bases 1 to 15088)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .15088
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            79. .1335
                     /gene="PA4142"
     CDS             79. .1335
                     /gene="PA4142"
                     /codon_start=1
                     /transl_table=11
                     /product="probable secretion protein"
                     /protein_id="AAG07529.1"
                     /db_xref="GI:9950348"
                     /translation="MFRQEALDAQHAGGLGEIVLIRPVSFTFLTLLAAAMALLVVGFF
                     LFGSYTKRSTVSGQLVPASGQVKVHAPQAGIVLRKFVQEGQAVRRGERLMVLSSERYG
                     SDAGPVQAGISRRLEQRRDSLRDELEKLRRLQDDERDSLTSKVASLQRELTTLAAQTD
                     SQQRLLALASDAAARYQGLMDKGYISMDQLQQRQAELLGQRQTLQGLERERTSLRQQL
                     TERRNELAGLSARQANQLAETRRQLSAVEQDLAESEAKRTLLVTAPESGIATAVLAEA
                     GQTVDSSRPLLSIVPADTPLQAELYAPSKSIGFIRPGDAVLIRYQAYPYQKFGQYHGK
                     VQSISRASVSYAELSSMVGGVPGLGQDGEQLYRLRVTLDDQAVTAYGQPRPLQSGMLL
                     DADILQDTRRLYEWVLEPLYSLTGKL"
     gene            1345. .3504
                     /gene="PA4143"
     CDS             1345. .3504
                     /gene="PA4143"
                     /codon_start=1
                     /transl_table=11
                     /product="probable toxin transporter"
                     /protein_id="AAG07530.1"
                     /db_xref="GI:9950349"
                     /translation="MAFLDALALRLGRRLPLVLQTEATECGLACLAMIAGYHGHHTGL
                     MELRRRFSVSLKGISLKQLIQTAHRLGLGTRAVKLDLGDLGKLKLPCVLHWNFNHFVV
                     LKAVDGRGAVLHDPAHGQRRLGLEEVSRSFTGVALELWPESGFEKQEAPPRIKLLGML
                     GKVTGLYRSLAQVLLLAGALEVFSLISPFFLQWTIDNVIVSEDRDLLSTLAIGFGLLL
                     LMQQAVSGVRAWVMMHMSTLLGVQWQANVFSHLLRLPAQYFEKRHLGDVVSRFGAVNS
                     IQQTLTAAFLSAVLDGLMTVATLGMMLLYSPPLAAIAIAAMSLYALGRWIWYRPLRNA
                     TEEQIVHAARQQSHFLETVRGIRPLKLFQRQDERRSVWLGLLVEQINAGLRTQKLQLF
                     YQQLNGLLFGVENLLVIWLGATMVMDGQFSVGILMAFNAYKSQFDSRVGSLIDKFFEL
                     RMLQLQGERLADIVLQAPEVSHGDILPENLRREASIEIQGLRYRYAEQEPWVLDGLD
                     LRIAGGESVAIVGPSGCGKSTLFNVLLGILPPVEGQIRMAGLDLAQLGLDGLRELVGT
                     VLQDDVLFAGSLSDNISFFDPQPDMPWLLQCAQMAAIHDDIQAMPMGYNTLVGDMGTV
                     LSGGQKQRVMLARALYKKPRILFLDEATSHLDVHCEQRVNAAIRALRITRIMVAHRPE
                     TIASADRVIVLGQGKVSLDESTARLAERQAAAAREQA"
     gene            3504. .4919
                     /gene="PA4144"
     CDS             3504. .4919
                     /gene="PA4144"
                     /codon_start=1
                     /transl_table=11
                     /product="probable secretion protein"
                     /protein_id="AAG07531.1"
                     /db_xref="GI:9950350"
                     /translation="MRALAGLLCGLLGLVPGAAAYEPDVFGTTGQVAGQAVYDLGGSG
```

```
                LPCRGGPPPTELSLEEAIERILCHDPQTRLAWANAKAQAAQVGIGKSAYLPRLDGRLD
                ASRGYSDMDYRDAPYLSGDGHRHRRGASLQLSWVLFDFGRRSAALRNAQQLLLAANAS
                QDATLQNTFALAAQAYYDALAAQRSLAASRQVAELAAQNLEAADAKYRAGAAALSDRL
                QAQTALSQASLAQVRDEGALSNALGVIALRMGLAPDTPLRLSGELEAQPDTGFVKAID
                EMLAEARREHPALLAAQARLKAAAASVEESRAAGRPSLALSANLARSHSDQAMAFNGD
                TRERDRSIGLQLNIPLFEGFERTYQVRNALARREASEAELADTEQQVSLEVWNNYQSL
                SVETRSLARTRELVEQSRQSLEVVQGRYRSGVGSMIELLNALTAYASAEDQHIRALGN
                WQTSRLRLAASLGRLGFWSLR"
gene            5015. .5905
                /gene="PA4145"
CDS             5015. .5905
                /gene="PA4145"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG07532.1"
                /db_xref="GI:9950351"
                /translation="MIDDIRYLIVFAKIAETGSVSGGAEALGLSAATASLHLSRLERN
                LGRALLYRNSRKLSLTQDGTQLLDTARSMLELYEKGFIEFRQRAVSTRNSLRISMPAV
                FVNGDFTRHLAAFIEEHPDLDLGIACDDSRHDIIGDGIDIAFRIGDLPDSSLKARHLF
                LLPRQVVAAPAFLARHAPLEHPRDLQRLEWIGLGMRPDSRLFRHARDDEEVRVDYTPR
                VRVDSVEASCRLARLGVGLAAPPTYLSDEPIRRGELRAVLPEWRLEPLKVYAVWPPNV
                PASSIAYTLINRLYEAFNPG"
gene            6093. .6497
                /gene="PA4146"
CDS             6093. .6497
                /gene="PA4146"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG07533.1"
                /db_xref="GI:9950352"
                /translation="MRARAGEQGHAGSAGDPGLGIEPLQEHALEEPGDPVFLQGQAWI
                RGGDAVAHMSDERRAGHPQVVEEARVGLDRADEPEADEGQQQDEAEGGAGDERQGPAE
                TMVGAGSDDGQVHRTGGNRQAQGERAHGEQQG"
gene            complement(6708. .8585)
                /gene="acoR"
                /note="PA4147"
CDS             complement(6708. .8585)
                /gene="acoR"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator AcoR"
                /protein_id="AAG07534.1"
                /db_xref="GI:9950353"
                /translation="MLSAHSKDHVDCVSRVLRNAARLPQAPVPTPILDSWRRSMEQHR
                LDPGSLQGPRILSQTLLDECRERSELFLRIASEEVARLHGRVRDADYCVLLTDAQGRT
                IDYRVESTIRNDCRKAGLYLGTCWSEGEEGTCGVAAVLTDRTAVTVHKRDHFRAAFID
                LTCSAAPVFDPQGELLGVLDVSAVQSPDDRRSQHLIRQMVAQSAREIENAFFMHSAQG
                HWVLRAHRAPGYVDSQPDYLFAWDADGRLLALNPAARHYLRQRLGQVPSHIGEVFDLD
                ALHAAADQAPCLLRGQAGELHVRLSAPRAKARRLEREVPDDAALDPRIAESLRLAVRV
                KDRNLPVLIQGETGAGKEVFARQLHQASARRDKPFVALNCAAIPESLIESELFGYVGG
                AFTGAAAKGMRGLLQQADGGTLFLDEIGDMPLGLQTRLLRVLAEGEVAPLGAARRQAV
                DIQVICATHRDLAALVAAGGFREDLYFRLGGARFELPPLRERSDRLALIRRILDEETA
                HCGVRIELGEAALECLLGYRWPGNVRQLRHVLRYACALCGGATLQLADLPAELRGERR
                TPASACESGGGPERDALLDALVRHRWKPMAAARELGISRATLYRRVRRHGIRMPGQAD
                D"
gene            8917. .9717
                /gene="PA4148"
CDS             8917. .9717
                /gene="PA4148"
                /codon_start=1
                /transl_table=11
                /product="probable short-chain dehydrogenase"
                /protein_id="AAG07535.1"
                /db_xref="GI:9950354"
                /translation="MHATSSDDFSLSSRVALVTGAGRGIGRGIALALARAGADVAVAD
                LDPQVAEETAAAIRSLGRRSLALGVDVSDGDSVRAMVERVATEFGRLDVAVNNAGVIS
                IRKVAELSLADWDRVMNVNARGVFLCCQAELPLMQAQRWGRIVNLSSIAGKVGLPDLA
                HYCASKFAVIGFSNALAKEVARDGVTVNALCPGIVGTGMWRGEDGLSGRWRQAGESEA
                QSWERHQASLLPQGEAQTVEDMGQLVVYLACAPHVTGQAIAVDGGFSL"
gene            9714. .10754
                /gene="PA4149"
CDS             9714. .10754
                /gene="PA4149"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG07536.1"
                /db_xref="GI:9950355"
                /translation="MTPPSPIVGIIANPASGRDLRRLTAHAGLHSSTDKASVVQRLLA
                AFGATGIRHALLPPDMTGIAAAVLRASQTRQARDGHWPFLEFLDMPLRQSVEDTRLAA
                RRMVERGVALIAVLGGDGTHKAVAAEVGDVPLLALSTGTNNAFPELREATGAGLAGGL
                CASNRVPPEIGLRRNKRLLVREPRRGLCEWALVDVAVSPQPFIGARAISRAEDLAEVF
                VSFAEPHAIGLSALCGLWSPVSREEPHGAWMRLHPDARESLLVPLAPGLLVGCGVSAA
                GYLQPGVAHAPSLSSGTLALDGEREIEFSATDRPSITLDPSGPFSVDVPATLAYAARH
                RLLAGQRTPMTP"
gene            10777. .11751
                /gene="PA4150"
CDS             10777. .11751
                /gene="PA4150"
                /codon_start=1
                /transl_table=11
                /product="probable dehydrogenase E1 component"
                /protein_id="AAG07537.1"
                /db_xref="GI:9950356"
                /translation="MSTLSTDQLLHAYRVMRTIRAFEERLHVEFATGEIPGFVHLYAG
                EEASAAGVMAHLRDDDCIASTHRGHGHCIAKGVDVHGMMAEIYGKKTGVCQGKGGSMH
                IADLEKGMLGANGIVGAGAPLAAGAALAAKLKGSDAVAVAFFGDGGSNEGAVFEAMNL
                ASVWNLPCLFVAENNGYAEATAANWSVACDHIADRAAGFGMPGVTVDGFDFFAVHEAA
                GAAIERARAGEGPSLIEVKLTRYYGHFEGDAQTYRDPDEVKHYRETRDCLKQFRERTC
                HAGLLSASDLDAIDAEVEARIEDAVQRAKNDPKPEPDDLLRDVYVSYP"
gene            11784. .12803
                /gene="acoB"
                /note="PA4151"
CDS             11784. .12803
                /gene="acoB"
                /codon_start=1
                /transl_table=11
                /product="acetoin catabolism protein AcoB"
                /protein_id="AAG07538.1"
                /db_xref="GI:9950357"
                /translation="MARKISYQQAINEALAQEMRRDPSVFIMGEDNAGGAGAPGEDDA
```

Query Match 5.6%; Score 63.8; DB 1; Length 15088;
Best Local Similarity 46.5%; Pred. No. 1.5e-06;
Matches 326; Conservative 0; Mismatches 357; Indels 18; Gaps 3;

```
Qy    432 CAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGA 491
          || |||||  || ||||| ||||||       |   |||   |   |  |||||
Db  14330 CAACGGCGCCTTCGCCGAATACGTCAACGTACCGGCCAACCTGCTCTACGCCCTGCCGGC 14389

Qy    492 CTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGT 551
          |    |||||  |||    || || ||| | |||||  |  | || | |  ||| || ||
Db  14390 CGGATTTCCCAGCGAGGCCGGCGCGCTGATCGAGCCGCTGGCGGTGGGCATGCACGCGGT 14449

Qy    552 TGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGG 611
            |   ||  |   || ||  || ||||  |   ||||  | || |  ||  | |||||
Db  14450 GAAGAAGGCCGGCAGCCTGCTCGGGCAGAACGTGGTGGTGGTCGGCGCCGGCACCATCGG 14509

Qy    612 CCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCC 671
          |||   ||||||  |     ||  | |    ||||   |  |  |       |||
Db  14510 CCTGAGCACCATCATGTGCGCGCGCGCCGCCGGCGCGGCACAGGTGATCGCCCTGGAGAT 14569

Qy    672 GGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTAC 731
          | ||| |    | |         ||   |||  | |||||      | |        | |
Db  14570 GTCCTCGGCGCGCAAGGCCAAGGCCCTGGAGGTCGGCGCCAGCCAGGTGCTCGACCCCTC 14629

Qy    732 ATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGC 791
            |  ||||||    | |   ||  ||     | | |         ||   | | |
Db  14630 GCGGTGCGACGCCCTAGGCGAAATCCGCGC---CCTCACCGGCGGGCTTGGCGCCGACGT 14686

Qy    792 AGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACC 851
            |||||| |||   ||       || ||  ||      | || |  ||||   |
Db  14687 CAGCTTCGAATGCATCGGCAACAAGCACACCGCCAAGCTGGCCATCGATGCCATCCGCAA 14746
```

```
Qy    852 TTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAAT 911
Db  14747 GGCCGGCAAATGCGTGCTTGTGGGAATCT---TCGAGGAGCCCAGTGAATTCAATTTCTT 14803

Qy    912 GTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTT 971
Db  14804 CGAGCTGGTTTCCACCGAGAAGCAGTTGCTCGGGGCCCTCGCCTACAACGGCGAGTTCGC 14863

Qy    972 CCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGAT 1031
Db  14864 CGACGTGATCGCCTTCATAGCCGACGGCCGCCT-----------CGACATCGCACCGCT 14911

Qy   1032 GATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGA 1091
Db  14912 GGTGACCGGTCGCATCGGCCTGGAAGAAATCGTCGAGCGCGGTTTCGAAGAGTTGGTGAA 14971

Qy   1092 GCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACG 1132
Db  14972 CAACAAGGAGCACAACGTGAAGATCATCGTCTCGCCGGGCG 15012
```

Search completed: May 2, 2003, 09:10:20
Job time : 3438 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 06:47:36 ; Search time 299 Seconds
(without alignments)
8608.810 Million cell updates/sec

Title: US-10-020-674-1
Perfect score: 1143
Sequence: 1 atgaaaggtttactttatta.....cgccgaacgaggtttcctaa 1143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	93.6	8.2	536	21	AAF08292	Fusarium venenatum
	2	83	7.3	1041	24	AAD28420	Bacillus subtilis
	3	69.8	6.1	339	24	ABN77781	Human dehydrogenas
	4	67	5.9	1113	21	AAC49213	Arabidopsis thalia
	5	67	5.9	1316	21	AAC49216	Arabidopsis thalia
	6	67	5.9	1347	21	AAC36132	Arabidopsis thalia
	7	67	5.9	1347	21	AAC49218	Arabidopsis thalia
	8	67	5.9	1413	21	AAC49219	Arabidopsis thalia
	9	65.4	5.7	1315	21	AAC38499	Arabidopsis thalia
	10	65.4	5.7	1407	21	AAC33156	Arabidopsis thalia
	11	63.8	5.6	1092	23	AAS54244	Pseudomonas aerugi
	12	62.8	5.5	1344	23	ABL03803	Drosophila melanog
	13	62.2	5.4	1338	23	ABL07985	Drosophila melanog
	14	57.4	5.0	933	22	AAH43317	T. aestivum orbito
	15	56.6	5.0	1163020	24	ABQ67197	Listeria innocua c
	16	56.6	5.0	3011208	24	ABQ69245	Listeria innocua D
	17	56.4	4.9	589	21	AAC57017	Pinus radiata tran
	18	56.4	4.9	589	21	AAA67351	Pinus radiata D-xy
	19	56.4	4.9	590	21	AAA67340	Pinus radiata D-xy
	20	55.4	4.8	1038	24	AAD33483	Human drug metabol
c	21	52.8	4.6	4164	23	ABL03802	Drosophila melanog
	22	52.2	4.6	2410	21	AAF15951	Human prostate can
	23	52.2	4.6	2519	24	ABN95691	Gene #2189 used to
	24	52.2	4.6	2519	24	ABL62665	Colon adenocarcino
	25	52.2	4.6	2519	24	ABL69415	Prostate cancer re
	26	52.2	4.6	2597	23	ABV23357	Human prostate exp
	27	52.2	4.6	2597	23	ABV25354	Human prostate exp
	28	52.2	4.6	2597	23	ABV29211	Human prostate exp
	29	51.8	4.5	1301	24	ABQ70384	Listeria monocytog
	30	51.4	4.5	534	21	AAA67346	Pinus radiata D-xy
	31	50.2	4.4	2944528	24	ABA03041	Listeria monocytog
	32	50	4.4	484	21	AAA67329	Eucalyptus grandis
	33	49.4	4.3	872	22	AAH52673	S. epidermidis ope
	34	49.4	4.3	1065	24	ABN91074	Staphylococcus epi
	35	49.4	4.3	4019	22	AAH54846	S. epidermidis gen
	36	49	4.3	1610	22	AAH43316	G. max sorbitol de
c	37	49	4.3	2703	23	AAS92993	DNA encoding novel
	38	47.2	4.1	2774	15	AAQ66579	Sorbitol dehydroge
	39	46.6	4.1	1302	24	ABQ68679	Listeria monocytog
	40	46	4.0	461	21	AAC56883	Pinus radiata tran
	41	45.8	4.0	1014	24	ABK73260	Bacillus lichenifo
c	42	45.8	4.0	3866	23	ABL07984	Drosophila melanog
	43	45.4	4.0	1053	21	AAF08294	Fusarium venenatum
	44	45	3.9	513	21	AAA77873	cDNA encoding huma
	45	45	3.9	513	22	AAI28611	Colon tumour relat

ALIGNMENTS

RESULT 1
AAF08292
ID AAF08292 standard; cDNA; 536 BP.
XX
AC AAF08292;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:815.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.

XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86; Page 695; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 536 BP; 126 A; 116 C; 162 G; 130 T; 2 other;

Query Match 8.2%; Score 93.6; DB 21; Length 536;
Best Local Similarity 51.6%; Pred. No. 3.9e-18;
Matches 213; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

Qy 371 CCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTG 430
Db 74 CGTGCAAGCGCGGCCTTCAGAACTGCTGTGATAAAAACGGTTTCATCGGGTTGAGTGGCT 133

Qy 431 CCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAG 490
Db 134 GGGGAGGTGGGTTGGCCGAATCAACTGTTGTGCCGCAGGATGCTGTCAAGAAGCTGCCCG 193

Qy 491 ACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTG 550
Db 194 ACAATGTGTCCCTAGAAGTCGGCGCTTTAGTCGAACCGCTTGCTGTTGGATGGCACGCAA 253

Qy 551 TTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCG 610
Db 254 TCAAGATCTCACCCTACGAGGAAGGCAACTCGGCTTTTGTTGTTGGTGGCGGACCAATTG 313

Qy 611 GCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGC 670
Db 314 GACTGGCTGTTGTGCAGGCTCTCATAGGACGTGGATGCAAGAACATCATGCTGAGCGAGG 373

Qy 671 CGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTA 730
Db 374 TCAGCTCCAAGAGACGCGAGTTTGCCAAGAAGTTTGGTGCGCATCACGTGTTCGATCCTA 433

Qy 731 CATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGA 783
Db 434 CAAANGACGATGTCGTGGCGAATGTTGAAAAAATTTGACAGGAGGGCTGGGAA 486

RESULT 2
AAD28420
ID AAD28420 standard; DNA; 1041 BP.
XX
AC AAD28420;
XX
DT 22-APR-2002 (first entry)
XX
DE Bacillus subtilis ydjL gene.
XX
KW Gene expression; cell metabolism; bioreactor health; fermentation;
KW genotypic alteration; ydjL gene; ds.
XX
OS Bacillus subtilis.
XX
PN WO200202766-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-US20873.
XX
PR 29-JUN-2000; 2000US-214967P.
PR 13-FEB-2001; 2001US-268320P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bedzyk LA, Wang T, Ye RW;
XX
DR WPI; 2002-154741/20.
XX
PT Expression of coding region of interest in a Bacillus species, for
PT monitoring fermentation process, involves growing transformed Bacillus
PT under oxygen depletion, the presence of nitrate or at various growth
PT stages -
XX
PS Claim 4; Page 44; 73pp; English.
XX
CC The invention relates to a method for expressing a coding region of
CC interest in a Bacillus sp. The method involves growing a transformed
CC Bacillus sp. having a chimeric gene comprising a nucleic acid fragment
CC comprising the promoter region of a Bacillus gene operably linked to a
CC coding region of interest expressible in Bacillus sp., under oxygen
CC limiting condition, in the presence of nitrate, or in the presence of
CC oxygen at various growth phases. The method is useful for regulating gene
CC expression in Bacillus sp. and for monitoring the cell metabolism of
CC Bacillus sp. culture. The method is useful for monitoring fermentation
CC and regulating bioreactor health. The method is also useful for
CC detecting genotypic alterations amount strains. The genes may be used in
CC modelling systems to test perturbation in fermentation process conditions
CC which will determine the requirements for the high yield of bioprocess
CC production. The present sequence is Bacillus subtilis ydjL gene used in
CC the invention.
XX
SQ Sequence 1041 BP; 304 A; 232 C; 263 G; 242 T; 0 other;

Query Match 7.3%; Score 83; DB 24; Length 1041;
Best Local Similarity 49.9%; Pred. No. 9.8e-15;
Matches 240; Conservative 0; Mismatches 235; Indels 6; Gaps 1;

Qy 407 TGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTG 466
Db 338 TGGGATTCCTCGGCTTAGCCGGCGGAGGCGGCGGTTTCTCTGAATACGTCTCTGTGGATG 397

Qy 467 AGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGC 526
Db 398 AAGAGCTTTTGTTCAAACTTCCTGATGAATTATCATATGAACAAGGCGCGCTCGTTGAAC 457

Qy 527 CTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCC 586
Db 458 CTTCTGCAGTTGCTCTATACGCTGTCCGCTCAAGCAAACTCAAAGCAGGCGACAAAGCGG 517

Qy 587 TGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATG 646
Db 518 CTGTATTCGGCTGCGGCCCGATCGGACTTCTTGTCATTGAAGCGCTGAAGGCTGCCGGTG 577

Qy 647 CGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGG 706
Db 578 CAACTGATATTTACGCTGTTGAGCTTTCTCCTGAACGCCAGCAAAAAGCTGAGGAGCTTG 637

```
Qy    707 GCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGG 766
          ||||     |  | ||||| |||| |   ||||| | |   ||||      ||  |   |
Db    638 GCGCGATCATCGTTGATCCGTCTAAAACAGACGATGTAGTCGCTG------AGATTGCAG 691

Qy    767 TGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCA 826
            |       |  ||  |  | | ||| |||||   | ||||||| ||   |   ||
Db    692 AACGTACAGGAGGCGGTGTTGACGTAGCATTCGAAGTCACTGGTGTCCCAGTGGTGTTAC 751

Qy    827 CCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAG 886
            | ||     ||||   | | | ||     ||||||  ||   | ||||||  |
Db    752 GACAAGCCATCCAGTCCACTACAATTGCCGGTGAAACCGTCATCGTCAGCATTTGGGAAA 811

Qy    887 A 887
          |
Db    812 A 812
```

RESULT 3
ABN77781
ID ABN77781 standard; cDNA; 339 BP.
XX
AC ABN77781;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human dehydrogenase-like ORF2728 cDNA, SEQ ID NO:5455.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US17076.
XX
PR 24-MAY-2000; 2000US-206690P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Leach MD, Shimkets RA;
XX
DR WPI; 2002-106200/14.
DR P-PSDB; ABP33755.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation -
XX
PS Claim 1; Page 1615; 2508pp; English.
XX
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
SQ Sequence 339 BP; 78 A; 70 C; 101 G; 90 T; 0 other;

Query Match 6.1%; Score 69.8; DB 24; Length 339;
Best Local Similarity 50.4%; Pred. No. 6e-11;
Matches 170; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

```
Qy    254 GCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAGCTACGTCGCATT 313
          || ||||||| | ||   |||||   || ||||  || |  || ||  | ||       |
Db      2 GCCCTGGTGTGAAAAACTTGAAAGTGGGAGACAAGGTAGTTGTCGAGCCCACAGGTACAT 61

Qy    314 GCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCT 373
          ||   ||| | |  || |          ||     |  |||   |     ||   || |
Db     62 GCAGAGACCGGTATCGTTGGCCCCTGTCGCCAAACGTTGACAAGGAATGGTGCGCTGCTT 121

Qy    374 GCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCA 433
          || | |  || |     |||   ||| | |   || | | || |||  |||||     |
Db    122 GCAAAAAGGGCTACTATAACATTTGTTCATATTTGGGGCTTTGTGGTGCGGGTGTGCAGA 181

Qy    434 GCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACT 493
          |||| || ||||| ||    || ||       ||    |||     |||  | || ||||
Db    182 GCGGTGGATTTGCAGAACGTGTTGTGATGAACGAATCTCACTGCTACAAAGTACCGGACT 241

Qy    494 CGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTG 553
            | |||   ||  ||| ||| || || | ||  |  ||||   |||||||||  ||
Db    242 TCGTGCCCTTAGACGTTGCAGCTTTGATTCAACCGTTGGCTGTGTGCTGGCATGCAATTA 301

Qy    554 AACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGT 590
          |  |      ||| |  | ||    |||||  || |
Db    302 GAGTCTGCGAGTTCAAAGCAGGCTCTACGGCTTTGAT 338
```

RESULT 4
AAC49213
ID AAC49213 standard; DNA; 1113 BP.
XX
AC AAC49213;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 60340.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 67; DB 21; Length 1113;
Best Local Similarity 47.1%; Pred. No. 8.2e-10;
Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

```
Qy    357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
Db    345 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 404

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
Db    405 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 464

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
Db    465 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 524

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
Db    525 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 584

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
Db    585 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 644

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
Db    645 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 704

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
Db    705 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 764
```

```
Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
Db    765 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 824

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
Db    825 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 871
```

RESULT 5
AAC49216
ID AAC49216 standard; DNA; 1316 BP.
XX
AC AAC49216;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 60351.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 67; DB 21; Length 1316;
Best Local Similarity 47.4%; Pred. No. 9e-10;
Matches 200; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

Qy 357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416

```
Db    433 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 492

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476

Db    493 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 552

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536

Db    553 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 612

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596

Db    613 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 672

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656

Db    673 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 732

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716

Db    733 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 792

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776

Db    793 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 852

Qy    777 CG 778

Db    853 AG 854
```

RESULT 6
AAC36132
ID AAC36132 standard; DNA; 1347 BP.
XX
AC AAC36132;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12661.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

```
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             5.9%;  Score 67;  DB 21;  Length 1347;
  Best Local Similarity   47.1%;  Pred. No. 9.1e-10;
  Matches  248;  Conservative    0;  Mismatches  270;  Indels    9;  Gaps    1;

Qy    357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
          | ||   || |    ||||  |   |||      |||   ||| |     |||  |||| 
Db    462 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTATGTCCAGAAATGAAGTTCTT 521

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
          |          |     |||  ||| || |  | || ||| |   || ||| |
Db    522 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAGGTGGTTCATCCTGCGGATCTATG 581

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
            ||||  |||| ||     |       ||    |||||| ||  ||| ||  ||  |||
Db    582 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 641

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
          ||     ||||||  |   || |||   '|    ||||     |  |   ||||  | ||
Db    642 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 701

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
          ||  || ||||| |||||||  || ||  | |||   || |       || |   | |||
Db    702 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCAGGCTTTCAGTGTGCCTAGAAT 761

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
          |||   |   ||    |    || | |   |    | |||||  | || |||||  || |
Db    762 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 821

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
          | | |  |     |||    ||   |   ||| || |   ||| |     |   ||||
Db    822 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 881

Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
            |                || ||| | | ||||||||  | ||| ||    | ||  |  |
Db    882 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 941

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
            ||  || | |  |||||  |   |   || |  | |     ||| |
Db    942 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 988


RESULT 7
AAC49218
ID   AAC49218 standard; DNA; 1347 BP.
XX
AC   AAC49218;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 60359.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
```

XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 67; DB 21; Length 1347;
Best Local Similarity 47.1%; Pred. No. 9.1e-10;
Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

```
Qy    357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
Db    464 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 523

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
Db    524 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 583

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
Db    584 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 643

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
Db    644 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 703

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
Db    704 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 763

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
Db    764 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 823

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
Db    824 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 883

Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
```

```
Db    884 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 943

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
Db    944 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 990
```

RESULT 8
AAC49219
ID AAC49219 standard; DNA; 1413 BP.
XX
AC AAC49219;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 60363.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 67; DB 21; Length 1413;
Best Local Similarity 47.1%; Pred. No. 9.3e-10;
Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

Qy 357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
Db 530 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTTTGTCCAGAAATGAAGTTCTT 589

```
Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
Db    590 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAAGTGGTACACCCTGCGGACCTATG 649

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
Db    650 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 709

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
Db    710 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 769

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
Db    770 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 829

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
Db    830 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 889

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
Db    890 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 949

Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
Db    950 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 1009

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
Db   1010 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 1056
```

RESULT 9
AAC38499
ID AAC38499 standard; DNA; 1315 BP.
XX
AC AAC38499;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21195.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.7%; Score 65.4; DB 21; Length 1315;
Best Local Similarity 47.2%; Pred. No. 2.8e-09;
Matches 199; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```
Qy    357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
          | ||   || |    ||||  |   |||      |||   ||| |     |||  ||||
Db    430 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTATGTCCAGAAATGAAGTTCTT 489

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
          |           |     |||  |||  ||  |  | || |||  |   || |||  |
Db    490 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAGGTGGTTCATCCTGCGGATCTATG 549

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
            ||||  |||| ||     |        ||    |||||| ||  ||| ||  ||  |||
Db    550 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 609

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
          ||     ||||||  |   || |||     |    ||||       |  |  ||||  | ||
Db    610 TGGTGTCCATGCTTGTCGCCGAGCTGAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 669

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
          ||  || ||||| |||||||  || ||  | |||   |  |       || |   | |||
Db    670 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 729

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
          |||   |   ||    |    || | |   |    | |||||  | || ||||| || |
Db    730 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 789

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
          | | |  |     |||    ||   |  ||| || |    |||  |     |   ||||
Db    790 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 849

Qy    777 CG 778
          |
Db    850 AG 851
```

RESULT 10
AAC33156
ID AAC33156 standard; DNA; 1407 BP.
XX
AC AAC33156;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 1991.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

```
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             5.7%;  Score 65.4;  DB 21;  Length 1407;
  Best Local Similarity   46.9%;  Pred. No. 2.9e-09;
  Matches  247;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    1;

Qy    357 TGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTG 416
          | ||   || |    ||||  |   |||      |||   ||| |     |||  ||||
Db    522 TTGGAGATGCAATCTCTGCAGGGAAGGACGATACAACCTATGTCCAGAAATGAAGTTCTT 581

Qy    417 CGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACAT 476
            |           |     |||  ||| ||  |  | || ||| |   || ||| |
Db    582 TGCAACTCCACCGGTTCATGGCTCTTTAGCTAACCAGGTGGTTCATCCTGCGGATCTATG 641

Qy    477 GGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGT 536
            ||||  |||| ||     |       ||    |||||| ||  ||| ||  ||  |||
Db    642 CTTCAAATTGCCTGAGAATGTGAGTTTGGAGGAAGGAGCAATGTGTGAACCACTTAGTGT 701

Qy    537 TGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGG 596
          ||     ||||||  |   || |||     |    ||||     |  |   ||||  | ||
Db    702 TGGTGTCCATGCTTGTCGCCGAGCTTAGGTTGGTCCTGAAACAAACGTTTTGGTAATGGG 761

Qy    597 AGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAAT 656
          ||  || ||||| |||||||  || ||  | |||   |  |       || |   | |||
Db    762 AGCTGGACCTATTGGCCTTGTTACAATGTTGGCTGCTCGGGCTTTCAGTGTGCCTAGAAT 821

Qy    657 TGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGT 716
          |||   |   ||    |   || | |    |     | |||||  | || |||||  || |
Db    822 TGTTATTGTGGATGTTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTCGGCGCAGATGA 881

Qy    717 GTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776
          | | |  |     |||    ||   |   ||| || |    |||  |     |    ||||
Db    882 GATTGTACAAGTGACAACAAACTTAGAGGATGTTGGTTCTGAGGTTGAACAGATTCAGAA 941

Qy    777 CG---------AGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
            |               || ||| | | ||||||||  | ||| ||    | ||  |  |
Db    942 AGCTATGGGGTCAAACATCGATGTAACATTCGACTGTGCGGGTTTTAACAAAACCATGTC 1001

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGG 874
          ||  || | | |||||  |   |   || |  | |     ||| |
Db   1002 GACAGCATTAGCAGCCACTCGTTGTGGCGGTAAAGTCTGTCTTGTCG 1048


RESULT 11
AAS54244
ID   AAS54244 standard; DNA; 1092 BP.
XX
AC   AAS54244;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Pseudomonas aeruginosa DNA for cellular proliferation protein #375.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU36385.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 7881; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1092 BP; 191 A; 380 C; 352 G; 169 T; 0 other;

  Query Match             5.6%;  Score 63.8;  DB 23;  Length 1092;
  Best Local Similarity   46.5%;  Pred. No. 7.8e-09;
  Matches  326;  Conservative    0;  Mismatches  357;  Indels   18;  Gaps    3;
```

```
Qy    432 CAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGA 491
          || |||||  || ||||| ||||||          |   |||   |   |   |||||
Db    405 CAACGGCGCCTTCGCCGAATACGTCAACGTACCGGCCAACCTGCTCTACGCCCTGCCGGC 464

Qy    492 CTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGT 551
          |    |||||  |||    || || ||| | |||||  |  | || | |  ||| || ||
Db    465 CGGATTTCCCAGCGAGGCCGGCGCGCTGATCGAGCCGCTGGCGGTGGGCATGCACGCGGT 524

Qy    552 TGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGG 611
          |   ||  |   || ||  || |||| |   |||| | || |  ||  | |||||
Db    525 GAAGAAGGCCGGCAGCCTGCTCGGGCAGAACGTGGTGGTGGTCGGCGCCGGCACCATCGG 584

Qy    612 CCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCC 671
          |||   ||||||  |      ||  | |      ||||   |  |  |      |||
Db    585 CCTGAGCACCATCATGTGCGCGCGCCGCCGGCGCGGCACAGGTGATCGCCCTGGAGAT 644

Qy    672 GGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTAC 731
          | ||| |   | |        ||   |||  | |||||     | |        |  |
Db    645 GTCCTCGGCGCGCAAGGCCAAGGCCCTGGAGGTCGGCGCCAGCCAGGTGCTCGACCCCTC 704

Qy    732 ATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGC 791
          |   ||||||    |  |   ||  ||      |  | |         ||   | | |
Db    705 GCGGTGCGACGCCCTAGGCGAAATCCGCGC---CCTCACCGGCGGGCTTGGCGCCGACGT 761

Qy    792 AGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACC 851
          ||||||  |||   ||       || ||   ||      | || |  ||||   |
Db    762 CAGCTTCGAATGCATCGGCAACAAGCACACCGCCAAGCTGGCCATCGATGCCATCCGCAA 821

Qy    852 TTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAAT 911
          | || |    |||   |||||   | |     ||  | || | || |||||      |
Db    822 GGCCGGCAAATGCGTGCTTGTGGGAATCT---TCGAGGAGCCCAGTGAATTCAATTTCTT 878

Qy    912 GTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTT 971
            |||  || |   |||||  |      |||  || |   ||||| || | ||  |
Db    879 CGAGCTGGTTTCCACCGAGAAGCAGTTGCTCGGGGCCCTCGCCTACAACGGCGAGTTCGC 938

Qy    972 CCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGAT 1031
          | | |   | | |         || |   | ||            | |         | |
Db    939 CGACGTGATCGCCTTCATAGCCGACGGCCGCCT-----------CGACATCGCACCGCT 986

Qy   1032 GATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGA 1091
          | | ||  ||     ||  |||  | ||    ||||||   ||  ||  ||  || || |  |
Db    987 GGTGACCGGTCGCATCGGCCTGGAAGAAATCGTCGAGCGCGGTTTCGAAGAGTTGGTGAA 1046

Qy   1092 GCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACG 1132
            |||||||| |||| || |||||| | ||  |||||  ||
Db   1047 CAACAAGGAGCACAACGTGAAGATCATCGTCTCGCCGGGCG 1087
```

RESULT 12
ABL03803
ID ABL03803 standard; cDNA; 1344 BP.
XX
AC ABL03803;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5891.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB59700.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 5891; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1344 BP; 319 A; 364 C; 367 G; 294 T; 0 other;

Query Match 5.5%; Score 62.8; DB 23; Length 1344;
Best Local Similarity 49.1%; Pred. No. 1.8e-08;
Matches 166; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

```
Qy    373 TGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCC 432
          ||| ||   ||      |||  ||  |   ||  |||||| |    |          |
Db    565 TGCAAGCAGGGCAAATACAACCTGTGCCCCGGAATGGTCTTCTGTGCCACGCCCCCCTAC 624

Qy    433 AGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGAC 492
            ||||    |  ||   |||  |    | |  | ||||  |    ||||||||| |||
Db    625 GACGGCAACCTCACCCGCTACTACAAGCATGCGGCGGACTTCTGCTTCAAGCTGCCCGAC 684

Qy    493 TCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTT 552
              |  ||   ||    || || ||| | |||||| | ||||| | |  ||||||
Db    685 CACGTCACCATGGAGGAGGGCGCCCTGCTCGAGCCTCTGTCTGTGGGCGTGCATGCCTGC 744

Qy    553 GAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGC 612
            | || ||     ||   |  ||    |  |  ||  | || |||| ||| || || |||
Db    745 AAGCGGGCGGAGGTCACCCTGGGCTCAAAAGTTCTCATCCTGGGAGCAGGACCCATTGGC 804

Qy    613 CTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCG 672
          || | |||  |  | |||   |||| |     ||||    | ||  |   | | || |
Db    805 CTGGTCACTTTAATGGCTGCTCAAGCCATGGGTGCGTCTGAGATCCTCATTACGGATCTT 864

Qy    673 GCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGC 710
          |    |    |     |  | || || || || || ||
Db    865 GTGCAGCAACGCCTCGATGTAGCTAAAGAGCTAGGTGC 902
```

RESULT 13
ABL07985
ID ABL07985 standard; cDNA; 1338 BP.
XX
AC ABL07985;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18437.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

```
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63882.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 18437; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1338 BP; 293 A; 360 C; 386 G; 299 T; 0 other;

  Query Match             5.4%;  Score 62.2;  DB 23;  Length 1338;
  Best Local Similarity   49.0%;  Pred. No. 2.7e-08;
  Matches  166;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy     372 CTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGC 431
Db     364 CTGCAAGCAGGGTCGATACAACCTGTGCGCCGACATGGTCTTTTGTGCCACACCTCCGTA 423

Qy     432 CAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGA 491
Db     424 CGATGGAAACCTGACCAGGTACTACAAACACGCCGCCGACTTTTGCTTCAAGCTGCCGGA 483

Qy     492 CTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGT 551
Db     484 TCACGTCAGCATGGAGGAAGGAGCTCTGCTGGAGCCCCTGTCCGTGGGAGTTCACGCCTG 543

Qy     552 TGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGG 611
Db     544 CCGTCGCGCAGGAGTCGGTCTGGGCTCAAAGGTGCTCATTCTGGGCGCTGGTCCAATTGG 603

Qy     612 CCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCC 671
Db     604 TTTGGTCACCCTGCTGGCTGCCCAAGCTATGGGTGCATCCGAGATCCTTATCACCGATCT 663

Qy     672 GGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGC 710
Db     664 GGTGCAGCAGCGTTTGGATGTTGCCAAGGAGCTGGGCGC 702


RESULT 14
AAH43317
ID   AAH43317 standard; cDNA; 933 BP.
XX
AC   AAH43317;
XX
DT   15-NOV-2001  (first entry)
XX
DE   T. aestivum orbitol dehydrogenase cDNA.
XX
KW   Aldehyde reductase; clustal alignment; plant specific promoter;
KW   carbohydrate metabolism; sorbitol biosynthesis; carbon partitioning;
KW   NADPH-dependant mannose 6-phosphate dehydrogenase;
KW   sorbitol dehydrogenase; ss.
XX
OS   Triticum aestivum.
XX
FH   Key             Location/Qualifiers
FT   CDS             29..931
FT                   /*tag= a
FT                   /product= "Sorbitol dehydrogenase"
FT                   /note= "No stop codon given"
XX
PN   US6274379-B1.
XX
PD   14-AUG-2001.
XX
PF   02-JUL-1999;   99US-0347803.
XX
PR   15-JUL-1998;   98US-0092952.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Famodu LO,  Kinney AJ,  Orozco EM;
XX
DR   WPI; 2001-556714/62.
DR   P-PSDB; AAB47474.
XX
PT   Nucleic acid sequence encoding a polypeptide having aldehyde reductase
PT   activity, useful for transforming plants with altered levels of
PT   expression of sorbitol biosynthetic enzymes -
XX
PS   Example 5; Column 53-54; 33pp; English.
XX
CC   The sequences given in AAH43311-17 encode enzymes which were used
CC   for comparison in the isolation of proteins having aldehyde reductase
CC   activity.  The isolated aldehyde reductase sequences were identified
CC   based on Clustal alignment method.  The aldehyde reductase cDNA,
CC   operatively linked to a plant specific promoter, is useful for
CC   altering the level of expression of aldehyde reductase in a
CC   transformed host cell. The aldehyde reductase is also useful in
CC   studies facilitating better understanding of carbohydrate metabolism
CC   and function in plants, for providing genetic tools for the manipulation
CC   of the sorbitol biosynthetic pathway, and for providing means to control
CC   carbon partitioning in plant cells.
XX
SQ   Sequence 933 BP; 159 A; 277 C; 350 G; 147 T; 0 other;

  Query Match             5.0%;  Score 57.4;  DB 22;  Length 933;
  Best Local Similarity   47.4%;  Pred. No. 6.6e-07;
  Matches  172;  Conservative    0;  Mismatches  191;  Indels    0;  Gaps    0;

Qy     359 GGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCG 418
Db     309 GGCGCTGCAGGCACTGCAAGGGCGGCCGCTACAACCTCTGCGACGACATGAAGTTCTTCG 368

Qy     419 GTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGG 478
Db     369 CCACCCCACCTTACCATGGATCACTTGCCGACCAGATTGTGCATCCAGGTGACCTGTGCT 428

Qy     479 TCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTG 538
Db     429 TCAAGCTTCCAGACAACGTGAGCCTGGAGGAGGGCGCCATGTGCGAGCCCCTGAGCGTGG 488

Qy     539 CCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAG 598
Db     489 GGGTGCACGCCTGCCGCCGAGCCGACGTGGGCGCGGAGAAGAGCGTGCTCATCATGGGCG 548

Qy     599 GAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTG 658
Db     549 CCGGCCCGATCGGCCTGGTCACCATGCTCTCGGCGCGCGCCTTCGGGGCGCCCAGGATCG 608
```

```
Qy     659 TGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGT 718
           |     ||||    | |     | |        | || |||   || ||||| || |
Db     609 TCATCGCCGACGTCGACGACCACCGCCTCTCCGTGGCCAAGTCCCTCGGCGCGGACGCCG 668

Qy     719 TCG 721
           |||
Db     669 TCG 671


RESULT 15
ABQ67197
ID   ABQ67197 standard; DNA; 1163020 BP.
XX
AC   ABQ67197;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria innocua contig DNA sequence #10.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria innocua.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators -
XX
PS   Claim 5; SEQ ID 10; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1163020 BP; 388339 A; 197589 C; 235612 G; 341474 T; 6 other;

  Query Match             5.0%;  Score 56.6;  DB 24;  Length 1163020;
  Best Local Similarity   48.3%;  Pred. No. 5.4e-05;
  Matches  158;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy     435 CGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTC 494
           ||| |  | ||| |||||  ||         |  ||  |  | || ||  | ||||||
Db  437437 CGGAGCCTATGCTGAGTATATCAAACTTCCAGCAGAAAATGTAGTAAAAATACCAGACGA 437496

Qy     495 GATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGA 554
             |    | ||    |  ||||| || ||||||  | | || | |   ||||   || |
Db  437497 ATTAGATTATGAAGCGGCTGCACTTGTCGAGCCTTCTGCAGTGGTCGTTCATGGATTTTA 437556

Qy     555 ACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCT 614
           |   | | |||  ||  ||||  |    | | | ||| | ||| | ||   |||||| |
Db  437557 TCATACGAAATTACAAGCTGGCGATGATGTCGTCGTTGTCGGAAGCGGCAATATCGGTTT 437616

Qy     615 TGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGC 674
             | ||||          |||    |   ||  | | | ||  |       ||    |
Db  437617 GCTAGCAATTCAATGGGCAAAAGTTTTTGGCGCAAGAAGAGTTATTGCGATTGATGTAGA 437676

Qy     675 CTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATG 734
           |   |  |    |   |||| || ||| |||| || || ||  |   |  ||||  |
Db  437677 CGATAAAAAGTTAGCTCTTGCGAAAGAAGTGGGAGCAGACGTAGTTATTAATTCTTTAAA 437736

Qy     735 TGACGACGCAAATGCTGTTCTCAAGGC 761
           || ||  ||   |  ||  ||   ||
Db  437737 AGAAGATCCATTAGAAGTAGTCGCAGC 437763


Search completed: May  2, 2003, 08:24:55
Job time : 1025 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 08:02:56 ; Search time 1611 Seconds
(without alignments)
11490.653 Million cell updates/sec

Title: US-10-020-674-1
Perfect score: 1143
Sequence: 1 atgaaaggtttactttatta.....cgccgaacgaggtttcctaa 1143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	735.2	64.3	1057	17	CNS0778D	AL432371 T3 end of
c	2	240.2	21.0	893	17	CNS07444	AL428330 clone BA0
c	3	233.4	20.4	946	17	CNS06YZB	AL421677 T3 end of
	4	195	17.1	1066	17	CNS06KBW	AL402690 T7 end of
	5	184	16.1	976	17	CNS06L93	AL403885 T7 end of
c	6	159.2	13.9	994	17	CNS06XQA	AL420056 T3 end of
	7	148.8	13.0	923	17	CNS073DD	AL427367 clone BA0
	8	120.4	10.5	908	17	CNS07443	AL428329 clone BA0
c	9	107.4	9.4	924	17	CNS070R5	AL423975 T3 end of
c	10	100.2	8.8	1045	17	CNS076OQ	AL431664 T7 end of
	11	99.8	8.7	982	17	CNS06L9E	AL403896 T7 end of
	12	97.6	8.5	1105	17	CNS07E9Y	AL441500 T3 end of
	13	94.4	8.3	489	17	AZ931833	AZ931833 474.dhz90
	14	77	6.7	493	9	AL696793	AL696793 AL696793
	15	62.6	5.5	596	12	BF717166	BF717166 Lf74 near
	16	62	5.4	673	13	BM513099	BM513099 kx72f07.y
	17	61.8	5.4	625	13	BM628275	BM628275.170006874
	18	60.8	5.3	698	13	BI574404	BI574404 RH23418.5
	19	60.2	5.3	432	12	BE756667	BE756667 210881 MA
	20	60.2	5.3	780	10	BE034697	BE034697 ML03B10 M
	21	59.4	5.2	492	12	BG366606	BG366606 HVSME1000
	22	59.4	5.2	508	10	AW465736	AW465736 BP230020A
	23	59.4	5.2	633	14	BQ472076	BQ472076 HV04F10r
	24	59.4	5.2	679	12	BF627089	BF627089 HVSMEb000
c	25	59.2	5.2	728	9	AA942405	AA942405 LD26593.3
	26	59	5.2	457	10	AW220586	AW220586 EST296971
	27	59	5.2	528	10	AW220587	AW220587 EST296972
	28	59	5.2	541	10	AW220588	AW220588 EST296973
	29	58.4	5.1	648	9	AI388895	AI388895 GH19951.5
	30	58	5.1	683	12	BG125422	BG125422 EST471068
	31	57.8	5.1	520	13	BM429385	BM429385 1A20E01 B
	32	57.8	5.1	561	12	BF618417	BF618417 HVSMEc000
	33	57.6	5.0	685	13	BI353054	BI353054 GM22035.5
	34	57.6	5.0	921	17	CNS078AH	AL433743 T7 end of
	35	57.2	5.0	671	13	BI627668	BI627668 RH69047.5
	36	57	5.0	467	12	BE756666	BE756666 210880 MA
	37	56.8	5.0	994	17	CNS078X5	AL434559 T7 end of
	38	56.4	4.9	653	13	BI624730	BI624730 RH64347.5
	39	56.4	4.9	664	13	BI616334	BI616334 RH46249.5
	40	56.2	4.9	672	9	AU001142	AU001142 AU001142
	41	55.8	4.9	449	10	AW282490	AW282490 LG1_313_A
	42	55.8	4.9	712	14	BQ805492	BQ805492 WHE3567_E
	43	55.8	4.9	720	14	BQ120882	BQ120882 EST606458
	44	55.6	4.9	488	10	BE215526	BE215526 HV_CEb000
	45	55.6	4.9	617	10	AV939921	AV939921 AV939921

ALIGNMENTS

RESULT 1
CNS0778D
LOCUS CNS0778D 1057 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone BB0AA009H12 of library BB0AA from strain CBS 4732 of Pichia angusta, genomic survey sequence.
ACCESSION AL432371
VERSION AL432371.1 GI:12215785
KEYWORDS GSS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE 1 (bases 1 to 1057)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1057)
AUTHORS Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta

```
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
   PUBMED   11152888
REFERENCE   3  (bases 1 to 1057)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1057
                     /organism="Pichia angusta"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA009H12"
                     /clone_lib="BB0AA"
                     /note="end : T3"
     misc_feature    <3. .>884
                     /note="similar to Saccharomyces cerevisiae ORF YAL061w [
                     FUN50 ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
     misc_feature    <45. .>799
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      249 a    259 c    283 g    263 t      3 others
ORIGIN

  Query Match           64.3%;  Score 735.2;  DB 17;  Length 1057;
  Best Local Similarity   94.7%;  Pred. No. 1.6e-208;
  Matches  780;  Conservative    2;  Mismatches   40;  Indels    2;  Gaps    2;

Qy    320 ACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCTGCCAGA 379
          ||||||||||||||| |||||||||||||| ||||| |||||||||||||||||| |||
Db      1 ACAGATCGCGCTACAAAGACACGGTCGCCCAGGACCTCGGGCTCTGTATGGCCTGCAAGA 60

Qy    380 GCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCG 439
          |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db     61 GCGGATCTCCAAACTGCTGTGTGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCG 120

Qy    440 GTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTC 499
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    121 GTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTTCCAGACTCGATTC 180

Qy    500 CCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCG 559
          |||||||||| |||||| |||||||||||||| ||||||||||||||||||||||||||
Db    181 CCGACGATATCGGAGCATTGGTTGAGCCTATTGCTGTTGCCTGGCATGCTGTTGAACGCG 240

Qy    560 CTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCA 619
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    241 CTAGATTCCAGCCTGGCCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCA 300

Qy    620 CCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGA 679
          |||||||||||||||||| | ||| ||||||||||||||||||||||||||||||||||
Db    301 CCATTCTTGCTCTGCAAGGCCACCGTGCCGGCAAAATTGTGTGTTCCGAGCCGGCCTTGA 360

Qy    680 TCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACG 739
          | |||||||||||||||||||||||||||||||||||||||| ||||||||||||||| |
Db    361 TTAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTTGATCCTTCTACATGTGATG 420

Qy    740 ACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCG 799
          |||||||||| |||||||||||||||||||||||| ||||| |||||||| ||||||||||

Db    421 ACGCAAATGCCGTTCTCAAGGCTATGGTGCCGGAAAACGAAGGATTCCACGCAGCCTTCG 480

Qy    800 ACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAA 859
          |||||||||| |||||||||||||||||||||||| |||||||||| || ||||| ||||
Db    481 ACTGCTCTGGAATTCCTCAGACATTCACCACCTCTATTGTCGCCACAGGCCCTTCGGGAA 540

Qy    860 TCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCTGA 919
          |||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||
Db    541 TCGCCGTCAACGTGGCCATTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCTGA 600

Qy    920 CTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGAAG 979
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTTACCAAGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGAAG 660

Qy    980 TTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGATGATTACAG 1039
          ||||| |||||||||||||||||||||||||||:||||| ||||||||||||||||||||
Db    661 TTGTCGGGGCCTTGGAAGATGGTCTCATATCTTTKGACAAGGCGCGCAAGATGATTACAG 720

Qy   1040 GCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGAGCACAAGG 1099
          |||||||||||||| |||||||| :|||||||||||||| |||||| ||||||||||||
Db    721 GCAAAGTCCACCTAAGGGACGGAG-YGAGAAGGGCTTTAGACAGCTCATCGAGCACAAGG 779

Qy   1100 AGAACAATGTCAAGATCCTG-GTGACGCCGAACGAGGTTTCCTA 1142
          | | |||||||||||||||| ||||| |||||||||||| |||||
Db    780 AAACCAATGTCAAGATCCTGTGTGACTCCGAACGAGGTGTCCTA 823


RESULT 2
CNS07444/c
LOCUS       CNS07444                 893 bp    DNA     linear   GSS 07-JUL-2001
DEFINITION  clone BA0AB025B04 of library BA0AB from strain CLIB 210 of
            Kluyveromyces lactis, genomic survey sequence.
ACCESSION   AL428330
VERSION     AL428330.1  GI:13432285
KEYWORDS    GSS.
SOURCE      Kluyveromyces lactis.
  ORGANISM  Kluyveromyces lactis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 893)
  AUTHORS   Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
            Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 11.
            Kluyveromyces lactis
  JOURNAL   FEBS Lett. 487 (1), 66-70 (2000)
  MEDLINE   20584721
   PUBMED   11152886
REFERENCE   3  (bases 1 to 893)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
```

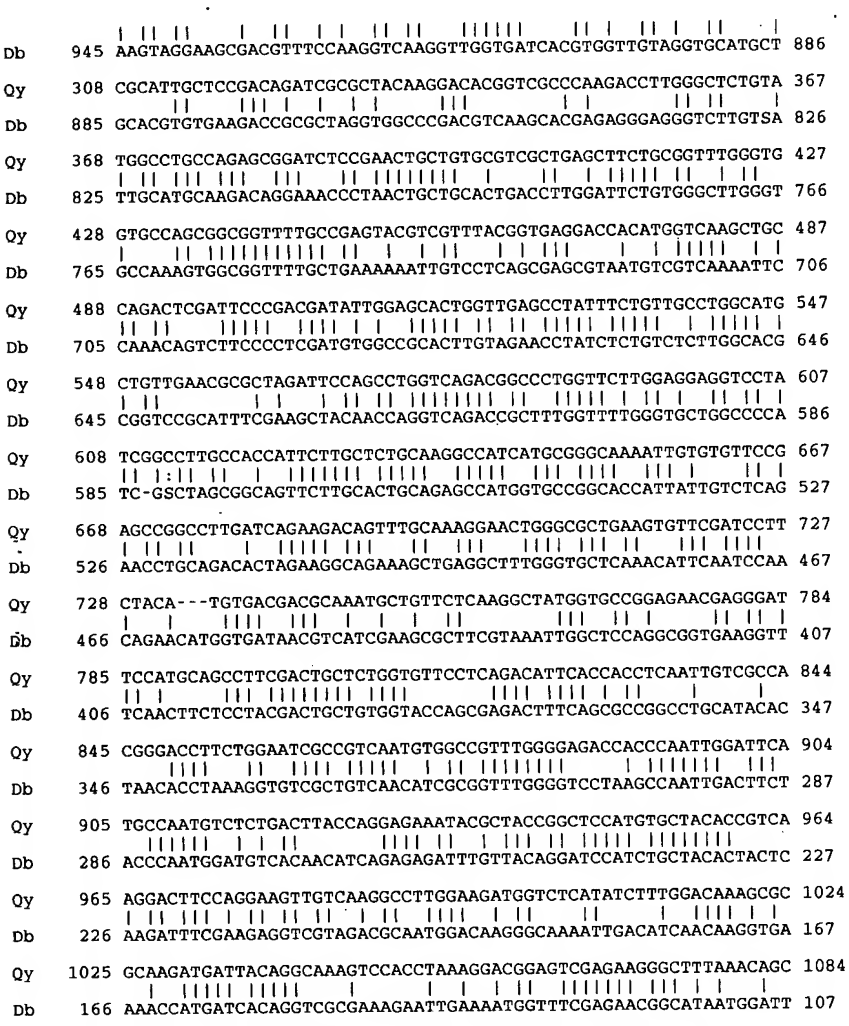
5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```
FEATURES             Location/Qualifiers
     source          1. .893
                     /organism="Kluyveromyces lactis"
                     /strain="CLIB 210"
                     /variety="lactis"
                     /db_xref="taxon:28985"
                     /clone="BA0AB025B04"
                     /clone_lib="BA0AB"
     misc_feature    complement(<182. .>850)
                     /note="similar to Saccharomyces cerevisiae ORF YAL061w [
                     FUN50 ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
     misc_feature    complement(<185. .>850)
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      261 a    177 c    177 g    278 t
ORIGIN
```

```
  Query Match             21.0%;  Score 240.2;  DB 17;  Length 893;
  Best Local Similarity   60.2%;  Pred. No. 2.8e-60;
  Matches  417;  Conservative    0;  Mismatches  273;  Indels    3;  Gaps    1;

Qy     441 TTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCC 500
           |||||| ||  | ||||||    ||||  | ||| | || ||| | ||| || |  | ||
Db     893 TTTTGCTGAAAAAGTCGTTGTTAGTGAAAAGCACGTTGTTAAGATTCCAAACACTTTACC 834

Qy     501 CGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGC 560
             ||  ||| |||  | || || || || || ||  | ||||||||||||  |  | |
Db     833 ATTGGACGTTGCAGCCTTAGTCGAACCAATCTCCGTCTCATGGCATGCTGTTAGAATCTC 774

Qy     561 TAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCAC 620
           |  || |||   || ||  | ||| | |||||||| | |||||| || ||  | |||||
Db     773 CAAGTTACAGAAGGGCCAATCCGCCTTAGTTCTTGGCGCAGGTCCAATTGGATTAGCCAC 714

Qy     621 CATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGAT 680
           |||| | ||| ||||||| |||  |||   ||| |||||    || || || ||    ||
Db     713 CATTTTAGCTTTGCAAGGTCATGGTGCCTCCAAGATTGTAGTCTCTGAACCAGCTGAAAT 654

Qy     681 CAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGT---GA 737
           |||||||  |    ||      | || || |  |||   ||||||||||||  |  |  ||
Db     653 CAGAAGAAATCAAGCTGCCAAGCTTGGTGTCGAAACCTTCGATCCTTCTGAACATAAAGA 594

Qy     738 CGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTT 797
           || ||   |  | || | |||    |||  || |     || || ||| ||   ||||
Db     593 AGATGCCGTTAATATTTTGAAGAAATTGGCTCCAGGTGGTGAAGGTTTCGATTTGCCTA 534

Qy     798 CGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGG 857
           ||| || |||||||||   |  || ||    ||      |    || ||          ||
Db     533 CGATTGTTCTGGTGTTAAACCAACTTTTGATACAGGTGTCCATGCTACTACCTTCAGAGG 474

Qy     858 AATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCT 917
           ||    || ||| | ||  | |||||  || | || || |  ||||  || |||    |
Db     473 TATGTATGTTAATATTGCAATCTGGGGCCACAAACCCATCGATTTCAAACCTATGGACGT 414

Qy     918 GACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGA 977
           |||   || ||||||| ||  ||||| |||||||| |||||| | || || ||  | ||
Db     413 CACTCTGCAAGAGAAATTCGTCACCGGTTCCATGTGTTACACCATTAAAGATTTTGAAGA 354

Qy     978 AGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGATGATTAC 1037
           || ||| | ||||| |   ||||   |||  || | || |||||  |  |  |||||||
Db     353 TGTGGTCCAAGCCTTAGGCAATGGCAGCATTGCTATTGATAAAGCTAGACATTTGATTAC 294

Qy    1038 AGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGAGCACAA 1097
           || | |    |  | | || ||  ||   || || ||  |  |  | ||  | || ||
Db     293 GGGTAGACAAAAAATCGAAGATGGGTTCACTAAAGGGTTCGACGAATTAATGAACCATAA 234

Qy    1098 GGAGAACAATGTCAAGATCCTGGTGACGCCGAA 1130
           ||||| ||  |||||||  || |||| || ||
Db     233 AGAGAAAAACATCAAGATATTGTTGACTCCTAA 201
```

```
RESULT 3
CNS06YZB/c
LOCUS       CNS06YZB                946 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T3 end of clone AY0AA014D05 of library AY0AA from strain CBS 6340
            of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION   AL421677
VERSION     AL421677.1  GI:12204873
KEYWORDS    GSS.
SOURCE      Kluyveromyces thermotolerans.
  ORGANISM  Kluyveromyces thermotolerans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 946)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 946)
  AUTHORS   Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 10.
            Kluyveromyces thermotolerans
  JOURNAL   FEBS Lett. 487 (1), 61-65 (2000)
  MEDLINE   20584720
   PUBMED   11152885
REFERENCE   3  (bases 1 to 946)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .946
                     /organism="Kluyveromyces thermotolerans"
                     /strain="CBS 6340"
                     /db_xref="taxon:4916"
                     /clone="AY0AA014D05"
                     /clone_lib="AY0AA"
                     /note="end : T3"
     misc_feature    complement(<45. .>883)
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]
                     1 putative frameshift(s)
                     similar to Saccharomyces cerevisiae ORF YAL061w [ FUN50 ;
                     similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      225 a    240 c    218 g    260 t      3 others
ORIGIN
```

```
  Query Match             20.4%;  Score 233.4;  DB 17;  Length 946;
  Best Local Similarity   55.3%;  Pred. No. 3.2e-58;
  Matches  492;  Conservative    1;  Mismatches  392;  Indels    4;  Gaps    2;

Qy     248 AGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAGCTACGT 307
```



RESULT 4
CNS06KBW
LOCUS CNS06KBW 1066 bp DNA linear GSS 17-JUN-2001
DEFINITION T7 end of clone AT0AA004A07 of library AT0AA from strain CBS 4311 of Saccharomyces servazzii, genomic survey sequence.
ACCESSION AL402690
VERSION AL402690.1 GI:12161979
KEYWORDS GSS.
SOURCE Saccharomyces servazzii.
ORGANISM Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1066)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1066)
AUTHORS Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H., Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE Genomic exploration of the hemiascomycetous yeasts: 7. Saccharomyces servazzii
JOURNAL FEBS Lett. 487 (1), 47-51 (2000)
MEDLINE 20584717
PUBMED 11152882
REFERENCE 3 (bases 1 to 1066)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .1066
/organism="Saccharomyces servazzii"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0AA004A07"
/clone_lib="AT0AA"
/note="end : T7"
misc_feature <2. .>967
/note="similar to Saccharomyces cerevisiae ORF YAL061w [FUN50 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental
BASE COUNT 323 a 253 c 233 g 252 t 5 others
ORIGIN

Query Match 17.1%; Score 195; DB 17; Length 1066;
Best Local Similarity 52.4%; Pred. No. 1.1e-46;
Matches 476; Conservative 2; Mismatches 422; Indels 9; Gaps 2;

```
Qy    534 TGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCT 593
          |||    |||||||||||| |    | | |||| |   |  |    | | | | || | ||
Db    301 TGTCAGTTGGCATGCTGTCGGTGTCTCCAGATACAGACAAGACGGGTCCGTCTTGATCCT 360

Qy    594 TGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAA 653
          || || || || ||||||||  |||| ||  | ||| || | || ||         | |
Db    361 CGGGGGTGGGCCAATCGGCCTAACCACAATCTTGGCTTTGAATGGACACGGGTGTACCGA 420

Qy    654 AATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGA 713
          || ||   ||| || || |||   || ||||||  |  | ||| |  |: ||||
Db    421 TATCGTCGTTTCAGAACCAGCCAAAATGAGAAGAGATCTAGCAGAAAARATGGGATGCCG 480

Qy    714 AGTGTTCGATCCTTCTA------CATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGT 767
          || ||||||||     |       |    || || | ||      | || |   | ||||
Db    481 TGTCTTCGATCCAAGCAAACACTCTGCAGAAGAAGGAACCAAGATCCTTAGATCAATGGC 540

Qy    768 GCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCAC 827
          || |      || || ||| |      | |||||||| || ||||||||    ||  | |
Db    541 TCCAGGCGGTGACGGGTTCGACTACTCGTTCGACTGTTCAGGTGTTCCAGCCACTCTTAG 600

Qy    828 CACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGG---G 884
          | || || |                 ||   ||| |||||||| ||| | ||||   |
Db    601 AGCTTCCATCGAATGTCTAACTTTCAGAGGTGCCGCTGTCAATGTCGCCATGTGGGACCG 660

Qy    885 AGACCACCCAATTGGATTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGG 944
          |   |  |  | |  ||| | ||||||  | | ||  | || || ||       |||||
Db    661 TGGGAAARCTGTCGATTTCTTCCCAATGGATATCACAAAACAAGAAAAGACTTACACCGG 720

Qy    945 CTCCATGTGCTACACCGTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCT 1004
          |||||||| ||||||    | || ||| | |||||  || | ||||| ||| | ||  |
Db    721 GTCCATGTGTTACACCTATCACGATTTCGAAGAAGTCATCGACGCCTTCGAAAAGGGCTT 780

Qy    1005 CATATCTTTGGACAAAGCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGT 1064
           ||   |  || | |||  || | ||||| | |  || |||| | |  |    |   |
Db    781 GATCGATCCAGAAAGAGCCAGCCACATGATAAGTGCAAAGGTCCCCTTGGAAAGAGCTTT 840

Qy    1065 CGAGAAGGGCTTTAAACAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCTGGTGAC 1124
           ||  | ||  | |      | |||   ||||| ||  | | | ||||  |  || | ||
Db    841 TGACGACGGTATCATGAGATTAATCACACACAAAGAAGAAACTATCAAAGTGTTGATCAC 900

Qy    1125 GCCGAACGA 1133
           || ||| |
Db    901 TCCAAACAA 909
```

```
RESULT 5
CNS06L93
LOCUS       CNS06L93                976 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AT0AA011E11 of library AT0AA from strain CBS 4311
            of Saccharomyces servazzii, genomic survey sequence.
ACCESSION   AL403885
VERSION     AL403885.1  GI:12164699
KEYWORDS    GSS.
SOURCE      Saccharomyces servazzii.
  ORGANISM  Saccharomyces servazzii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 976)
  AUTHORS   Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 7.
            Saccharomyces servazzii
  JOURNAL   FEBS Lett. 487 (1), 47-51 (2000)
  MEDLINE   20584717
   PUBMED   11152882
REFERENCE   3  (bases 1 to 976)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Saccharomyces servazzii"
                     /strain="CBS 4311"
                     /db_xref="taxon:27293"
                     /clone="AT0AA011E11"
                     /clone_lib="AT0AA"
                     /note="end : T7"
     misc_feature    <1. .>972
                     /note="similar to Saccharomyces cerevisiae ORF YAL061w [
                     FUN50  ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
     misc_feature    <7. .>969
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49  ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      291 a    151 c    240 g    270 t     24 others
ORIGIN
```

```
  Query Match             16.1%;  Score 184;  DB 17;  Length 976;
  Best Local Similarity   50.8%;  Pred. No. 2e-43;
  Matches  479;  Conservative   13;  Mismatches  443;  Indels    7;  Gaps    3;

Qy    180 CAAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAAC 239
          ||   | ||| |    | | | :|||||| |    || ||||||  ||  | || ||
Db      3 CAGACATAAGCTAAGTGAAAATRAACTTCCCCAGGCTCTTGGACACGAGGTGAGTGGTTA 62

Qy    240 GGTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGA 299
          |||    |  ||||  | |  || ||     | |||   || ||| ||||||  || ||
Db     63 TGTGAAAGCWATTGGACCAGCCGTGACTCAATTWAAAGTAGGGGACCGAGTCGTCGTCGA 122

Qy    300 AGCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGG 359
          ||| |   |  |:|    || |     || |       | |          ||
Db    123 AGCAAGTGCTAGTWGTGTTGATAAGCATCGTTGGCCCAATGCCAAAWWTGCTAACAMACC 182

Qy    360 GCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGG 419
          | |||   || ||  |   |||:     ||||| |||| |      | ||| | ||
Db    183 TATATGTGATGCTTGTAATGAAGGAWACGAAAACTGTTGTGAATATGCTGGTTTCAGTGG 242

Qy    420 TTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGT 479
          ||||||||  |     || || ||||| ||     | |    |  |||  | ||  | ||
Db    243 TTTGGGTGTAGTTTCTGGAGGGTTTGCAGACACGATGGAAACCATTGAACATCATGTCGT 302

Qy    480 CAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGC 539
          .  ||  : || || |   | ||    ||| |||| ||  | || |||||  | ||||||||
Db    303 TAAAWWACCTGATTTCTTACCTATGGATGTTGGTGCTTTAGTGGAGCCATTGTCTGTTGC 362

Qy    540 CTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGG 599
          |||||||  |  |       ||:|  ||     |||||| |||| ||  |  ::|||||  |
Db    363 ATGGCATGGTTCTCGTGTAGCWAATTTTACCCCTGGTAAGACTGCTTTAAWWCTTGGTGC 422

Qy    600 AGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGT 659
```

```
Db    423 AGGTCCAATTGGGCWAGCCATGAWCCTGGTGTTAAAGGCCAAARGTGCTAAGAAGATTGT 482

Qy    660 GTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTT 719

Db    483 WGTTAGTGAACTGGCATCTATCAGAAGAGAATTGGCTGAAAGAWTCAATGTGGAAACATT 542

Qy    720 CGATCCTTCTA---CATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAA 776

Db    543 TGAWCCTTCAAAGCATGGAGATAATRCWATTACCGAATWACGTTCGAACACAGAAGGTAA 602

Qy    777 CGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAAT 836

Db    603 TAAAGGGTTTGATTTTGCGTTTGACTGTACTGGTGTAGCACCAACTATCAATACAGGTAT 662

Qy    837 TGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAAT 896

Db    663 TGCTGCTATWCATTATAGAGGTACTTATTGTAACGTTGCCATTTGGGGCAAAGGATTAAA 722

Qy    897 TGGATTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTA 956

Db    723 T---TTTAACCCAATGGATATTACATTACAAGAGAAGAATTTGACTGGTTCTATTGGTTA 779

Qy    957 CACCGTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGA 1016

Db    780 TACAGTAGAAGATTTTAAGCAAGTTGTTGATGCATTTGAAAAGAAGAAGATCGATCCTAA 839

Qy   1017 CAAAGCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTT 1076

Db    840 AGAGTGTGAACATTTGATCACCGGTAGACAAAAATTGAA-GATGGTTGGGAAAAGGGGTT 898

Qy   1077 TAAACAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCT 1118

Db    899 CTTGGAGTTGATGAACCATAAGGATACAAACATCAAGGTKCT 940
```

RESULT 6
CNS06XQA/c
LOCUS CNS06XQA 994 bp DNA linear GSS 06-JUL-2001
DEFINITION T3 end of clone AY0AA003A03 of library AY0AA from strain CBS 6340 of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION AL420056
VERSION AL420056.1 GI:12203234
KEYWORDS GSS.
SOURCE Kluyveromyces thermotolerans.
ORGANISM Kluyveromyces thermotolerans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1 (bases 1 to 994)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 994)
AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 10. Kluyveromyces thermotolerans
JOURNAL FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE 20584720
PUBMED 11152885
REFERENCE 3 (bases 1 to 994)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .994
/organism="Kluyveromyces thermotolerans"
/strain="CBS 6340"
/db_xref="taxon:4916"
/clone="AY0AA003A03"
/clone_lib="AY0AA"
/note="end : T3"
misc_feature complement(<6. .>671)
/note="similar to Saccharomyces cerevisiae ORF YAL060w [FUN49 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental
misc_feature complement(<66. .>671)
/note="similar to Saccharomyces cerevisiae ORF YAL061w [FUN50 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental
BASE COUNT 243 a 263 c 242 g 244 t 2 others
ORIGIN

Query Match 13.9%; Score 159.2; DB 17; Length 994;
Best Local Similarity 54.4%; Pred. No. 5.6e-36;
Matches 365; Conservative 0; Mismatches 299; Indels 7; Gaps 2;

```
Qy      1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60

Db    671 ATGAGAGCGTTAGCATACTTTGGTCAAAAAGACATNCGCTTCACTGACAGCCTTGCGCAA 612

Qy     61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120

Db    611 CCACAGGTTTCTCATGAGGATGAGGTGGAAATTGACGTTAGCTGGTGCGGTGTATGTGGT 552

Qy    121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180

Db    551 TCCGATCTCCACGA------ATATCTTGATGGCCCTATTTTCTTTCCAGCTAATGGGAAG 498

Qy    181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240

Db    497 ACACACGCTATGAGCGGCCAAGGCCTTCCCCAGGCCATGGGCCATGAGATGTCTGGCATT 438

Qy    241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300

Db    437 GTGTCAAAAGTAGGAAGCGACGTTTCCAAGGTCAAGGTTGGTGATCACGTGGTTGTAGGT 378

Qy    301 GCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGG 360

Db    377 GCATGCTGCACGTGTGAAGACCGCGCTAGGTGGCCCGACGTCAAGCACGAGAGGGAGGGT 318

Qy    361 CTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGT 420

Db    317 CTGTGCATTGCATGCAAGACAGGAAACCCTAACTGCTGCACTGACCTTGGATTCTGTGGG 258

Qy    421 TTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTC 480

Db    257 CTTGGGTGCCAAAGTGGCGGTTTTGCTGAAAAAATTGTCCTCAGCGAGCGTAATGTCGTC 198

Qy    481 AAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCC 540

Db    197 AAAATTCCAAACAGTCTTCCCCTCGATGTGGCCGCACTTGTAGAACCTATCTCTGTCTCT 138

Qy    541 TGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGA 600

Db    137 TGGCACGCGGTCCGCATTTCGAAGCTACAACCAGGTCAGACCGCTTTGGTTTTGGGTGCT 78
```

```
Qy    601 GGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTG 660
          || || ||| |||| ||  |  ||||||| |||||  ||| ||||  ||| |
Db     77 GGCCCCATC-GCCTAGCGGCAGTTCTTGCACTGCAGAGCCATGGTGCCGGCACCATTATT 19

Qy    661 TGTTCCGAGCC 671
          || || ||
Db     18 GTCTCAGAACC 8
```

RESULT 7
CNS073DD
LOCUS CNS073DD 923 bp DNA linear GSS 07-JUL-2001
DEFINITION clone BA0AB018G03 of library BA0AB from strain CLIB 210 of Kluyveromyces lactis, genomic survey sequence.
ACCESSION AL427367
VERSION AL427367.1 GI:12210561
KEYWORDS GSS.
SOURCE Kluyveromyces lactis.
ORGANISM Kluyveromyces lactis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1 (bases 1 to 923)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 923)
AUTHORS Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F., Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R., Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
TITLE Genomic exploration of the hemiascomycetous yeasts: 11. Kluyveromyces lactis
JOURNAL FEBS Lett. 487 (1), 66-70 (2000)
MEDLINE 20584721
PUBMED 11152886
REFERENCE 3 (bases 1 to 923)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .923
/organism="Kluyveromyces lactis"
/strain="CLIB 210"
/variety="lactis"
/db_xref="taxon:28985"
/clone="BA0AB018G03"
/clone_lib="BA0AB"
misc_feature <152. .>820
/note="similar to Saccharomyces cerevisiae ORF YAL060w [FUN49 ; similarity to alcohol/sorbitol dehydrogenase] similar to Saccharomyces cerevisiae ORF YAL061w [FUN50 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental

```
BASE COUNT      252 a    160 c    232 g    268 t     11 others
ORIGIN

  Query Match           13.0%;  Score 148.8;  DB 17;  Length 923;
  Best Local Similarity  52.6%;  Pred. No. 6.9e-33;
  Matches  367;  Conservative    7;  Mismatches  306;  Indels   18;  Gaps    2;

Qy      9 TTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAACCGGAGAT 68
          ||: | |||||  || || |  ||||| || | |:||       | ||||||||  |  ||
Db    145 TTWGCGTTATTTTGGAACCAGAGATATCCGTTTCWCCACTGATCTACCTGAACCAGTAAT 204

Qy     69 CAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGCACGGACTT 128
            | || ||  | |||||  |||| |  |    |  ||||| |||||||| || ||  |
Db    205 TGAAAAACCGGATGATGTTTTGATCGATATAGTGTGGTGTGGTATCTGTGGTACTGATCT 264

Qy    129 GAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACCAAGGACAA 188
            | ||| ||    |       || ||    || ||||| ||| | ||    |   |  |
Db    265 ACATGAACTCCTCGAC------GGACCAAACTTCTTCCCAAAAGATGGTGAAACACATGA 318

Qy    189 GATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGA 248
          :|||   ||    | | | ||||         || || ||  |    ||  |||     |
Db    319 DATTAGTGGTAGGGGATTGCCTCAAGCCATGGGCCACGAGATGGCAGGCGTGGTTTCACA 378

Qy    249 GGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAGCTACGTC 308
            || |||||| |||| || |   |  |||| ||||||   || |  || ||  ||||
Db    379 AGTCGGCTCTAGTGTGACTAAATTCCAACCAGGTGACCATGTTGTTGTGGAGCCTACAGG 438

Qy    309 GCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGGCTCTGTAT 368
            ||    ||||||    | || ||   : ||             || :   || |
Db    439 TACATGTCAGGACAGAGAAAGGTATAATACTWAGG-----------GTGADACGTGCAA 486

Qy    369 GGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGG 428
            || ||  | || ||    |  ||   ||||  |    |  | || ||||| |  ||||
Db    487 AGCTTGTAAAAGGGGTATGCACAATGTTTGTGTGCATTTWGGGTTATGCGGATGTGGTGT 546

Qy    429 TGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCC 488
              || || ||||| ||  | || |||  ||||||| : ||      ||||  | ||
Db    547 ACAATCAGGTGGATTTGCTGAAAAAGTTGTTATCGGTGAGAWACATTGTTTCAAAGTACC 606

Qy    489 AGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGC 548
          |   | ||| ||    |:| | | |||| || || ||||| |  | |||  |||||||||
Db    607 ATCATGGATCCCACTTGWTGTGGCAGCATTGATTCAGCCTTTAGCAGTTTGCTGGCATGC 666

Qy    549 TGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTAT 608
          ||| ||   |     |   |||||  | ||  || |||| || || ||||| ||
Db    667 TGTGAAATTGGGAGCTGTGAAAGAAGGTCAATCTGCATTGATTCTCGGTGGTGGTCCGAT 726

Qy    609 CGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGA 668
            ||  | | ||  ||  | ||| |||||  |  |           | || |||  ||| ||
Db    727 TGGGTTGGGCATTATAATGGCTTTGCAACACTTTGGGTGCACTGATATCGTGGTTTCGGA 786

Qy    669 GCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGG 706
          ||| |||   || ||| |  |  | ||  |  ||  ||
Db    787 GCCTGCCAAAATTAGACGTGATCTAGCTGAAAAATGGG 824
```

RESULT 8
CNS07443
LOCUS CNS07443 908 bp DNA linear GSS 07-JUL-2001
DEFINITION clone BA0AB025B04 of library BA0AB from strain CLIB 210 of Kluyveromyces lactis, genomic survey sequence.
ACCESSION AL428329
VERSION AL428329.1 GI:13449456
KEYWORDS GSS.
SOURCE Kluyveromyces lactis.
ORGANISM Kluyveromyces lactis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1 (bases 1 to 908)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,

```
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 908)
  AUTHORS   Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
            Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 11.
            Kluyveromyces lactis
  JOURNAL   FEBS Lett. 487 (1), 66-70 (2000)
  MEDLINE   20584721
   PUBMED   11152886
REFERENCE   3  (bases 1 to 908)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .908
                     /organism="Kluyveromyces lactis"
                     /strain="CLIB 210"
                     /variety="lactis"
                     /db_xref="taxon:28985"
                     /clone="BA0AB025B04"
                     /clone_lib="BA0AB"
     misc_feature    <51. .>571
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]
                     similar to Saccharomyces cerevisiae ORF YAL061w [ FUN50 ;
                     similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      298 a    160 c    191 g    259 t
ORIGIN

  Query Match             10.5%;  Score 120.4;  DB 17;  Length 908;
  Best Local Similarity   53.3%;  Pred. No. 2.2e-24;
  Matches  304;  Conservative    0;  Mismatches  256;  Indels   10;  Gaps    2;

Qy     575 GTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGC 634
           ||||  | ||  || |||| || || ||||| || ||  | | ||  ||  | ||| |||
Db       1 GTCAATCTGCATTGATTCTCGGTGGTGGTCCGATTGGGTTGGGCATTATAATGGCTTTGC 60

Qy     635 AAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTG 694
           ||  |  |          | || |||  ||| ||||| |||  || ||| |  |  | |
Db      61 AACACTTTGGGTGCACTGATATCGTGGTTTCGGAGCCTGCCAAAATTAGACGTGATCTAG 120

Qy     695 CAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTC 754
           |  |  || |||| |||   || | | | |||      |   |   |   | |
Db     121 CTGAAAAAATGGGTGCTATTGTTTCCAACCCTATGGATTTCAAGACCAATGAGGAAAACA 180

Qy     755 TCAAGG---------CTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCT 805
           ||||           |  |    ||  |    || |||| | ||    | || || || |
Db     181 TCAAAAACTTATGTTCATTATCTCCAAATGGTGAAGGATACGATTTCACATTTGATTGTT 240

Qy     806 CTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCG 865
           | ||| || ||   || |||    | |||||                        ||||  |||
Db     241 CGGGTATTTCTGTCACTTTCGAATCATCAATCAAATGTCTAACTTTCAGAGGAACTGCCA 300

Qy     866 TCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCTGACTTACC 925
           ||||||| || || ||||| |||  | |  |          ||||||  | | || |  |
Db     301 TCAATGTTGCTGTGTGGGGTGACAGCTCTGTGAAGCATTATCCAATGGATTTAACATCAC 360

Qy     926 AGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGAAGTTGTCA 985
           |||| ||| |   ||| || || ||||| |||||  |  |||| ||  |||| || ||
Db     361 AGGAAAAAAAATATACAGGGTCAATGTGTTACACATTGGAGGATTTTGAGGACGTCGTTG 420

Qy     986 AGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGATGATTACAGGCAAAG 1045
           |  ||||    ||  ||   | |  |||  | ||||||      |||||  || || ||||
Db     421 ATTCCTTCACCGA-AGTAAAACAGATTTCAAAAAAGCGGCTTTCATGATCACGGGAAAAG 479

Qy    1046 TCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGAGCACAAGGAGAACA 1105
           |    ||| || | ||| | ||  | |  || |      |||||  |  |||| ||    |
Db     480 TTTCACTAGAGAATGGAATGGAAGACGCATTCATGAGATTGATCAATGACAAAGAAGGAA 539

Qy    1106 ATGTCAAGATCCTGGTGACGCCGAACGAGG 1135
              ||||||||||  | || || || || |
Db     540 CCATCAAGATCCTACTTACTCCAAAGGAAG 569


RESULT 9
CNS070R5/c
LOCUS       CNS070R5                 924 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T3 end of clone AZ0AA011F06 of library AZ0AA from strain CBS 712 of
            Kluyveromyces marxianus, genomic survey sequence.
ACCESSION   AL423975
VERSION     AL423975.1  GI:12207169
KEYWORDS    GSS.
SOURCE      Kluyveromyces marxianus.
  ORGANISM  Kluyveromyces marxianus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 924)
  AUTHORS   Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 12.
            Kluyveromyces marxianus var. marxianus
  JOURNAL   FEBS Lett. 487 (1), 71-75 (2000)
  MEDLINE   20584722
   PUBMED   11152887
REFERENCE   3  (bases 1 to 924)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Kluyveromyces marxianus"
                     /strain="CBS 712"
```

```
                     /variety="marxianus"
                     /db_xref="taxon:4911"
                     /clone="AZ0AA011F06"
                     /clone_lib="AZ0AA"
                     /note="end : T3"
     misc_feature    complement(<1. .>456)
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]
                     similar to Saccharomyces cerevisiae ORF YAL061w [ FUN50 ;
                     similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      290 a    239 c    139 g    254 t      2 others
ORIGIN

  Query Match             9.4%;  Score 107.4;  DB 17;  Length 924;
  Best Local Similarity   53.8%;  Pred. No. 1.7e-20;
  Matches  247;  Conservative    0;  Mismatches  206;  Indels    6;  Gaps    1;

Qy       1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60
           |||   |  |||   |||| |||||  || ||||||   |   |  |      |    ||
Db     456 ATGCGTGCGTTAGCGTATTTCGGTAAGAAAGATATTAAATTTACAAATGATTTAAAGGAG 397

Qy      61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120
           ||    ||   |     | | || ||  | || || ||    |  ||||| || |||||
Db     396 CCTACCATTGGGGTGGACGATGAAGTAGAAATTAACGTTGCTTGGTGTGGTATTTGTGGT 337

Qy     121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180
           |  ||| |  | |||| |  | ||       |||||  |||| |||||  || | || |
Db     336 AGTGACCTTCATGAATACCTAGAT------GGTCCAATTTTCTTCCCCGAAGATGGGAAG 283

Qy     181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
           |   |  |  |    ||    |   | ||||    |   || |||||| |   ||| |
Db     282 ACCCATGATGTCAGTGGTCTTGGTTTACCTCAGGCTATGGGGCATGAAATGTCCGGTATA 223

Qy     241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
           |||    ||||||| |||   || | ||  || || |||||||||| | || |  || |||
Db     222 GTGTCAAAGGTTGGATCTAAAGTGAAAAACGTCAAGCCTGGTGACCGTGTTGTTGTCGAA 163

Qy     301 GCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGG 360
           |||||   |  || |  ||    |  || |   |||   | |       | |   ||||
Db     162 GCTACTGGGTCATGTTTAGATCATTACCGTTGGGAGGGTTCTGCTCATGCAAAAGATGGG 103

Qy     361 CTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGT 420
           |||   || ||    || |||| |   ||||| |||||     |  | |||   |||
Db     102 AATTGTGCTGCTTGTAGTAGAGGATATTACAACTGTTGTGCCCATTTAGGGTTCATGGGT 43

Qy     421 TTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTT 459
           | ||||    |||||| |||||||| ||| | || |||
Db      42 CTTGGTGTGCACAGCGGTGGTTTTGCAGAGAAAGTTGTT 4


RESULT 10
CNS0760Q/c
LOCUS       CNS0760Q                1045 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BB0AA005F01 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL431664
VERSION     AL431664.1  GI:12215078
KEYWORDS    GSS.
SOURCE      Pichia angusta.
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 1045)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1045)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
   PUBMED   11152888
REFERENCE   3  (bases 1 to 1045)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the other
            extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1045
                     /organism="Pichia angusta"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA005F01"
                     /clone_lib="BB0AA"
                     /note="end : T7"
     misc_feature    complement(<876. .>977)
                     /note="similar to Saccharomyces cerevisiae ORF YAL060w [
                     FUN49 ; similarity to alcohol/sorbitol dehydrogenase ]"
                     /evidence=not_experimental
BASE COUNT      285 a    248 c    227 g    280 t      5 others
ORIGIN

  Query Match             8.8%;  Score 100.2;  DB 17;  Length 1045;
  Best Local Similarity   83.0%;  Pred. No. 2.8e-18;
  Matches  137;  Conservative    1;  Mismatches   24;  Indels    3;  Gaps    2;

Qy     978 AGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGATGATTAC 1037
           ||  ||   || | | | |    |||||||||||||    ||||||||||||  |:||
Db    1029 AGAAGTKTCGGGCCTTGGAAGATGTCTCATATCTTTG--ACAAGCGCGCAAGATATTWAC 972

Qy    1038 AGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGAGCACAA 1097
           | |||||||||||||| |||||||||||||||||||||| |||||| ||||||||||
Db     971 A-GCAAAGTCCACCTAAGGGACGGAGTCGAGAAGGGCTTTAGACAGCTCATCGAGCACAA 913

Qy    1098 GGAGAACAATGTCAAGATCCTGGTGACGCCGAACGAGGTTTCCTA 1142
           ||| | |||||||||||||||||||| |||||||||| |||||
Db     912 GGAAACCAATGTCAAGATCCTGGTGACTCCGAACGAGGTGTCCTA 868


RESULT 11
CNS06L9E
LOCUS       CNS06L9E                982 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AT0AA011F10 of library AT0AA from strain CBS 4311
            of Saccharomyces servazzii, genomic survey sequence.
ACCESSION   AL403896
VERSION     AL403896.1  GI:12164726
KEYWORDS    GSS.
SOURCE      Saccharomyces servazzii.
  ORGANISM  Saccharomyces servazzii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
```

de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 982)
AUTHORS Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H., Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE Genomic exploration of the hemiascomycetous yeasts: 7. Saccharomyces servazzii
JOURNAL FEBS Lett. 487 (1), 47-51 (2000)
MEDLINE 20584717
PUBMED 11152882
REFERENCE 3 (bases 1 to 982)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1..982
/organism="Saccharomyces servazzii"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0AA011F10"
/clone_lib="AT0AA"
/note="end : T7"
misc_feature <2..>538
/note="similar to Saccharomyces cerevisiae ORF YAL061w [FUN50 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental
BASE COUNT 329 a 195 c 178 g 277 t 3 others
ORIGIN

Query Match 8.7%; Score 99.8; DB 17; Length 982;
Best Local Similarity 53.4%; Pred. No. 3.5e-18;
Matches 256; Conservative 1; Mismatches 213; Indels 9; Gaps 2;

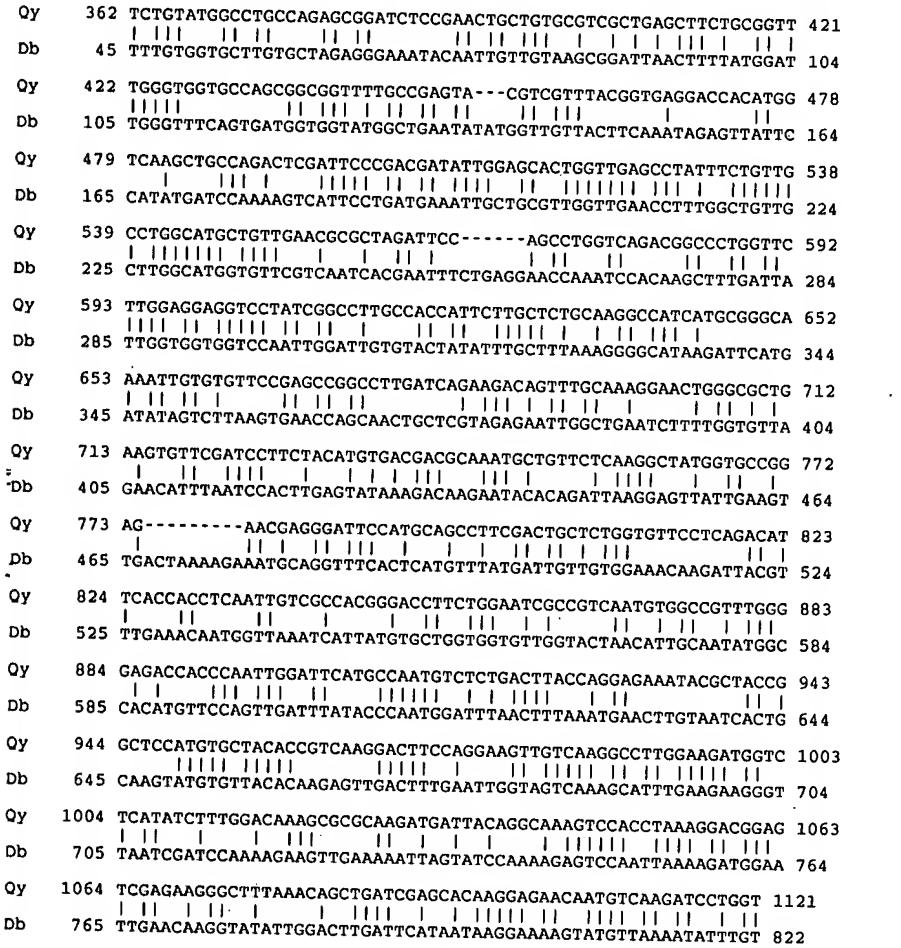
```
Qy    664 TCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGAT 723
          || || || |||   |: |||||| |  | ||| |  |  ||||       || ||||||
Db      2 TCAGAACCAGCCAAAAWGAGAAGAGATCTAGCAGAAAAGATGGGATGCCGTGTCTTCGAT 61

Qy    724 CCTTCTA------CATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAAC 777
          ||  |       |    || || | ||      | || |   | ||||  || |
Db     62 CCAAGCAAACACTCTGCAGAAGAAGGAACCAAGATCCTTAGATCAATGGCTCCAGGCGGT 121

Qy    778 GAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATT 837
          || || ||| |    | |||||||| || |||||||||   ||  | |   | || ||
Db    122 GACGGGTTCGACTACTCGTTCGACTGTTCAGGTGTTCCAGCCACTCTTAGAGCTTCCATC 181

Qy    838 GTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGG---GAGACCACCCA 894
          |                ||   ||| |||||||| ||| | ||||   | |   |  |
Db    182 GAATGTCTAACTTTCAGAGGTGCCGCTGTCAATGTCGCCATGTGGGACCGTGGGAAAGCT 241

Qy    895 ATTGGATTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGC 954
           | |  ||| | ||||||  | | ||  | || || ||       |||| ||||||||
Db    242 GTCGATTTCTTCCCAATGGATATCACAAAACAAGAAAAGACTTACACCGGGTCCATGTGT 301

Qy    955 TACACCGTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTG 1014
          |||||||    | || ||| | |||||  || | ||||| ||| | ||  | ||   |
Db    302 TACACCTATCACGATTTCGAAGAAGTCATCGACGCCTTCGAAAAGGGCTTGATCGATCCA 361

Qy   1015 GACAAAGCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGC 1074
          || | |||  || | ||||| | |  || |||| | |  |     |   | ||  | ||
Db    362 GAAAGAGCCAGCCACATGATAAGTGCAAAGGTCCCCTTGGAAAGAGCTTTTGACGACGGT 421

Qy   1075 TTTAAACAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACGA 1133
            | |       | |||   ||||| ||  | | | ||||  |  || | || || ||| |
Db    422 ATCATGAGATTAATCACACACAAAGAAGAAACTATCAAAGTGTTGATCACTCCAAACAA 480
```

RESULT 12
CNS07E9Y
LOCUS CNS07E9Y 1105 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone XBD0AA002D06 of library XBD0AA from strain CBS 94 of Candida tropicalis, genomic survey sequence.
ACCESSION AL441500
VERSION AL441500.1 GI:12224726
KEYWORDS GSS.
SOURCE Candida tropicalis.
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1105)
AUTHORS Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis
JOURNAL FEBS Lett. 487 (1), 91-94 (2000)
MEDLINE 20584726
PUBMED 11152891
REFERENCE 3 (bases 1 to 1105)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web :
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1..1105
/organism="Candida tropicalis"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone="XBD0AA002D06"
/clone_lib="XBD0AA"
/note="end : T3"
misc_feature <35..>829
/note="similar to Saccharomyces cerevisiae ORF YAL061w [FUN50 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental
misc_feature <47..>829

/note="similar to Saccharomyces cerevisiae ORF YAL060w [FUN49 ; similarity to alcohol/sorbitol dehydrogenase]"
/evidence=not_experimental

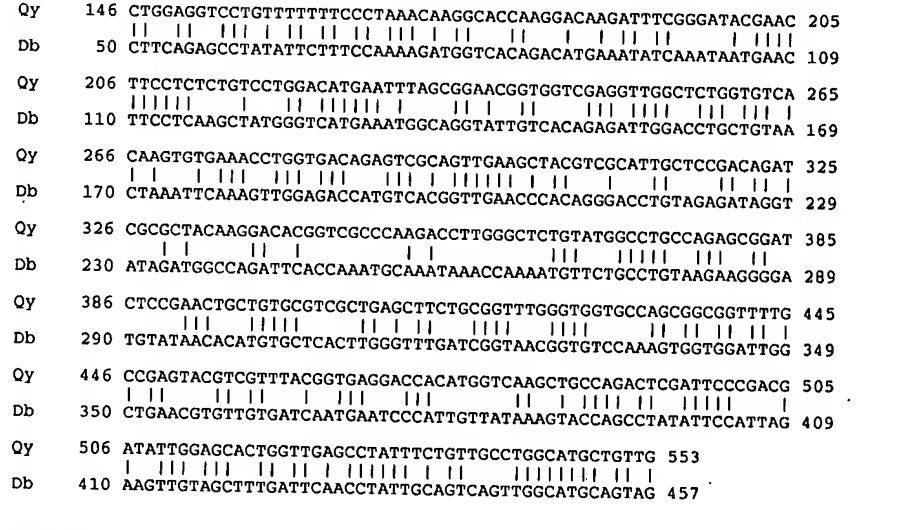
BASE COUNT 343 a 164 c 222 g 374 t 2 others
ORIGIN

Query Match 8.5%; Score 97.6; DB 17; Length 1105;
Best Local Similarity 48.3%; Pred. No. 1.7e-17;
Matches 376; Conservative 0; Mismatches 384; Indels 18; Gaps 3;



RESULT 13
AZ931833
LOCUS AZ931833 489 bp DNA linear GSS 01-APR-2001
DEFINITION 474.dhz90f10.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces unisporus genomic clone 474.dhz90f10.s1, DNA sequence.
ACCESSION AZ931833
VERSION AZ931833.1 GI:13502745
KEYWORDS GSS.
SOURCE Saccharomyces unisporus.
ORGANISM Saccharomyces unisporus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 489)
AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish ,W.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
FEATURES Location/Qualifiers
source 1. .489
/organism="Saccharomyces unisporus"
/strain="NRRL Y-1556 (CBS 398)"
/db_xref="taxon:27294"
/clone="474.dhz90f10.s1"
/clone_lib="Saccharomyces unisporus NRRL Y-1556"
/note="Random genomic sequence "
BASE COUNT 149 a 86 c 116 g 138 t
ORIGIN

Query Match 8.3%; Score 94.4; DB 17; Length 489;
Best Local Similarity 52.0%; Pred. No. 8.1e-17;
Matches 212; Conservative 0; Mismatches 196; Indels 0; Gaps 0;



RESULT 14
AL696793
LOCUS AL696793 493 bp mRNA linear EST 21-MAR-2002
DEFINITION AL696793 NAP1 Anopheles gambiae cDNA clone NAP1-P46-B-01-5, mRNA sequence.
ACCESSION AL696793
VERSION AL696793.1 GI:19616315
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

```
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Christophides,G.K., Blass,K., Zdobnov,E.M., Carmouche,R., Benes,V.
            and Kafatos,F.C.
  TITLE     Anopheles gambiae EST, European Molecular Biology Laboratory
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Christophides GK
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany
            Tel: +49 6221 387-440
            Fax: +49 6221 387-306
            Email: christop@embl-heidelberg.de
            Plate: P46 row: B column: 01.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Anopheles gambiae"
                     /db_xref="taxon:7165"
                     /clone="NAP1-P46-B-01-5"
                     /clone_lib="NAP1"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia); Site_1: NotI;
                     Site_2: EcoRI; ESTs sequenced from the T7 priming site
                     that reads from the 5' end of cDNA. The NAP1 is a
                     directionally cloned and normalized, oligo-T primed cDNA
                     library constructed from a mixture of Anopheles gambiae
                     developmental stages according to: Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery, Genome Research
                     6, 791-806."
BASE COUNT       89 a    128 c    167 g    107 t      2 others
ORIGIN

 Query Match             6.7%;  Score 77;  DB 9;  Length 493;
 Best Local Similarity   53.3%;  Pred. No. 1.3e-11;
  Matches  184;  Conservative    0;  Mismatches  160;  Indels    1;  Gaps    1;

Qy     372 CTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGC 431
           |||| || |||| ||    |||  |||||||  | ||| ||||||
Db     133 CTGCAAGGGCGGTTCGTACAACCTCTGTGCGGAGATGATCTTCTGTGCCACCCCGCCGTA 192

Qy     432 CAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGA 491
           |   |||    || ||  |||| | |    |    | |||         ||| |||||
Db     193 CGATGGCAACCTTACCCGGTACTTTGCCCATCCGGCCGACTTTTGCTACAAACTGCCCGA 252

Qy     492 CTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGT 551
           |    |  ||   ||    || || ||| | |||||  |||| || |    ||||||
Db     253 CCACGTGACCATGGAGGAGGGTGCCCTGCTGGAGCCGCTTTCGGTCGGTGTGCATGCGTG 312

Qy     552 TGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGG 611
           |   || || |    |   ||  |||  |  ||  ||  ||||||| |  || || |||||
Db     313 TCGGCGGGCAAACGTTGGGCTCGGTTCGCAGGTGCTAATTCTTGGGGCGGGACCGATCGG 372

Qy     612 CCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCC 671
            || |  ||  |  | |      |||       |||||||||| |  |   | |||| |
Db     373 TCTGGTGACGTTGATAGTGGCCAAAGCGATGGGTGCGGGCAAAGTGCTCGTTACCGATCT 432

Qy     672 GGCCTTGATCAGAAGACAGTTTGC-AAAGGAACTGGGCGCTGAAG 715
           |  |  || | |     |  | ||  ||||| ||||||||||| |
Db     433 GCTCCAGAACCGGCTGGACGTGGCGCAAGGAGCTGGGCGCTGATG 477


RESULT 15
BF717166
LOCUS       BF717166                596 bp    mRNA    linear   EST 12-DEC-2001
DEFINITION  Lf74 near-ripe peach fruit cDNA library Prunus persica cDNA clone
            Lf74 similar to sorbitol dehydrogenase (AB016256), mRNA sequence.
ACCESSION   BF717166
VERSION     BF717166.1  GI:17531052
KEYWORDS    EST.
SOURCE      peach.
  ORGANISM  Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Callahan,A.M., Richie,A.J., Byrne,G.B., Artlip,T.S. and Bassett
            ,C.L.
  TITLE     Expressed Sequence Tags of Fruit from 'Loring' Peach
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Ann Callahan
            USDA-ARS, Appalachian Fruit Research Station
            45 Wiltshire Road, Kearneysville, WV 25430, USA
            Tel: 304-725-3451
            Fax: 304-728-2340
            Email: acallaha@afrs.ars.usda.gov
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Prunus persica"
                     /cultivar="Loring"
                     /db_xref="taxon:3760"
                     /clone="Lf74"
                     /clone_lib="near-ripe peach fruit cDNA library"
                     /tissue_type="Whole fruit minus the stone"
                     /dev_stage="21-60 Newtons flesh resistance"
                     /note="Vector: Lambda vector UniZap XR; Site_1: EcoRI;
                     Site_2: XhoI; Fruit cDNA. The mRNA was isolated from whole
                     fruit from which the stone was excised. The fruit was near
                     ripe, registering from 21 to 60 newtons resistance to a
                     penetrometer. The library was constructed using the lambda
                     UniZap XR by Stratagene."
BASE COUNT      135 a    139 c    177 g    144 t      1 others
ORIGIN

 Query Match             5.5%;  Score 62.6;  DB 12;  Length 596;
 Best Local Similarity   48.5%;  Pred. No. 3.2e-07;
 Matches  173;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy     359 GGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCG 418
           ||  |||  |  ||| || ||||       ||   |||  |     |||  |||| ||
Db     181 GGAGATGTGAGCAGTGCAAGGGCGGGCGATACAATCTCTGCCCTGACATGAAGTTCTTCG 240

Qy     419 GTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGG 478
                 | |  ||||   || ||  |  | | ||  ||  ||  || |
Db     241 CCACCCCACCAGTCCACGGCTCATTGGCAAATCAGATTGTGCACCCTGCAGATCTGTGCT 300

Qy     479 TCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTG 538
           |||||||  || ||      |     ||    || ||  ||   ||||||  |   |||||
Db     301 TCAAGCTCCCGGAGAATGTGAGTTTGGAAGAAGGCGCCATGTGCGAGCCTCTGAGTGTTG 360

Qy     539 CCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAG 598
           |||||  |  || || |  ||  || |    |  | ||||||  | ||||
Db     361 GTGTTCATGCCTGTCGGCGAGCCAACATCGGCCCAGAAACAAATGTCCTGGTGATCGGAG 420

Qy     599 GAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTG 658
           ||| || ||||| ||||   |  | ||  ||   |  |    |  |||| || |||||
Db     421 CAGGGCCAATCGGGCTTGTGTCGGTGCTCTCTGCTCGTGCTTTTGGGGCGGCCAGAATTG 480

Qy     659 TGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAG 715
           |   |  ||  ||         |  |   ||||||||  || ||||| || |
Db     481 TCATTGTGGATGTGGATGATGAGCGCTTATCCATTGCAAAGTCTCTCGGCGCCGACG 537


Search completed: May  2, 2003, 09:32:06
Job time : 1626 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 08:04:46 ; Search time 100 Seconds
(without alignments)
3505.315 Million cell updates/sec

Title: US-10-020-674-1
Perfect score: 1143
Sequence: 1 atgaaaggtttactttatta.....cgccgaacgaggtttcctaa 1143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	57.4	5.0	933	4	US-09-347-803-21	Sequence 21, Appl
	2	49.4	4.3	1065	4	US-09-134-001C-537	Sequence 537, App
	3	49	4.3	1610	4	US-09-347-803-19	Sequence 19, Appl
	4	47.2	4.1	2774	3	US-08-466-548B-1	Sequence 1, Appli
	5	47.2	4.1	2774	3	US-07-998-226F-1	Sequence 1, Appli
	6	47.2	4.1	2774	5	PCT-US93-12560-1	Sequence 1, Appli
	7	45	3.9	1393	1	US-08-174-467-18	Sequence 18, Appl
	8	45	3.9	1393	3	US-08-452-071-18	Sequence 18, Appl
	9	43.4	3.8	1419	1	US-08-174-467-17	Sequence 17, Appl
	10	43.4	3.8	1419	3	US-08-452-071-17	Sequence 17, Appl
	11	41.8	3.7	1154	4	US-09-347-803-15	Sequence 15, Appl
	12	40.6	3.6	1283	1	US-08-174-467-19	Sequence 19, Appl
	13	40.6	3.6	1283	3	US-08-452-071-19	Sequence 19, Appl
c	14	39.8	3.5	30001	1	US-08-125-468-1	Sequence 1, Appli
c	15	39.8	3.5	30001	2	US-08-474-933-1	Sequence 1, Appli
	16	35.8	3.1	6877	1	US-08-347-340-1	Sequence 1, Appli
	17	35.6	3.1	1188	4	US-08-860-656B-9	Sequence 9, Appli
	18	35.6	3.1	1481	4	US-08-860-656B-1	Sequence 1, Appli
	19	35.4	3.1	1377	1	US-08-174-467-20	Sequence 20, Appl
	20	35.4	3.1	1377	3	US-08-452-071-20	Sequence 20, Appl
c	21	35.2	3.1	2061	4	US-09-800-960-1	Sequence 1, Appli
	22	33.2	2.9	2318	3	US-08-826-964-1	Sequence 1, Appli
	23	33	2.9	535	2	US-08-713-000-1	Sequence 1, Appli
	24	33	2.9	535	2	US-08-975-316-1	Sequence 1, Appli
	25	33	2.9	535	4	US-09-211-710-1	Sequence 1, Appli
	26	33	2.9	535	4	US-09-615-192A-1	Sequence 1, Appli
	27	33	2.9	578	4	US-09-615-192A-112	Sequence 112, App
	28	33	2.9	592	2	US-08-975-316-30	Sequence 30, Appl
	29	33	2.9	592	4	US-09-615-192A-30	Sequence 30, Appl
	30	33	2.9	1087	4	US-09-372-422A-29	Sequence 29, Appl
c	31	32.6	2.9	910	4	US-08-456-200B-16	Sequence 16, Appl
	32	32	2.8	1035	1	US-08-891-254-8	Sequence 8, Appli
	33	32	2.8	1035	2	US-08-819-539-8	Sequence 8, Appli
	34	32	2.8	1035	2	US-09-030-270A-8	Sequence 8, Appli
	35	32	2.8	1035	4	US-08-984-207-8	Sequence 8, Appli
	36	32	2.8	1035	4	US-09-013-587-8	Sequence 8, Appli
	37	32	2.8	1035	5	PCT-US96-08819-8	Sequence 8, Appli
	38	32	2.8	2277	1	US-08-676-967-2	Sequence 2, Appli
	39	32	2.8	2277	1	US-08-676-974-2	Sequence 2, Appli
	40	32	2.8	2277	2	US-09-098-487-2	Sequence 2, Appli
c	41	32	2.8	7791	2	US-08-149-097D-23	Sequence 23, Appl
c	42	32	2.8	7791	3	US-08-949-386-23	Sequence 23, Appl
c	43	32	2.8	7791	3	US-08-450-562-23	Sequence 23, Appl
c	44	32	2.8	7791	4	US-08-984-709A-23	Sequence 23, Appl
c	45	32	2.8	7791	4	US-08-450-272-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-347-803-21
; Sequence 21, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Orozco, Buddy
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-347-803-21

Query Match 5.0%; Score 57.4; DB 4; Length 933;
Best Local Similarity 47.4%; Pred. No. 1.9e-08;
Matches 172; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

```
Qy     359 GGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCG 418
           ||| ||| | |  |||| || ||||        |||  ||| |       |||  |||| ||
Db     309 GGCGCTGCAGGCACTGCAAGGGCGGCCGCTACAACCTCTGCGACGACATGAAGTTCTTCG 368

Qy     419 GTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGG 478
                 |  |    ||      |||||||  |  | ||  |     |  ||||
Db     369 CCACCCCACCTTACCATGGATCACTTGCCGACCAGATTGTGCATCCAGGTGACCTGTGCT 428

Qy     479 TCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTG 538
           |||||||  ||||||     |   |   ||    || ||  ||   |||||  |    || |
Db     429 TCAAGCTTCCAGACAACGTGAGCCTGGAGGAGGGCGCCATGTGCGAGCCCCTGAGCGTGG 488

Qy     539 CCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAG 598
             |||  ||        || ||      |     | |   |||  |  ||  |  | || |
Db     489 GGGTGCACGCCTGCCGCCGAGCCGACGTGGGCGCGGAGAAGAGCGTGCTCATCATGGGCG 548

Qy     599 GAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTG 658
             || || |||||||| | ||||||  ||   |   ||  | |       |||  ||  || |
Db     549 CCGGCCCGATCGGCCTGGTCACCATGCTCTCGGCGCGCGCCTTCGGGGCGCCCAGGATCG 608

Qy     659 TGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGT 718
           |     ||||    | |     | |             | || |||   || |||||  || |
Db     609 TCATCGCCGACGTCGACGACCACCGCCTCTCCGTGGCCAAGTCCCTCGGCGCGGACGCCG 668
```

```
Qy     719 TCG 721
           |||
Db     669 TCG 671


RESULT 2
US-09-134-001C-537
; Sequence 537, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 537
;   LENGTH: 1065
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-537

 Query Match             4.3%;  Score 49.4;  DB 4;  Length 1065;
  Best Local Similarity   45.8%;  Pred. No. 7.9e-06;
  Matches  213;  Conservative    0;  Mismatches  246;  Indels    6;  Gaps    1;

Qy     427 GGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTG 486
           ||| |    || |  ||||||||||        |     |||      |  |      | ||
Db     367 GGTTCTGATGGTGCATTTGCCGAGTTTACTAATGCTCCTGAAACAAATGTTTATCATCTA 426

Qy     487 CCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCAT 546
           |||||     || | |          || || || || || || || |   | |||||    ||
Db     427 CCAGATAATGTTTCAGCACGAGAAGGTGCTCTTGTAGAACCAACAGCCGTTGCTGTCCAA 486

Qy     547 GCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCT 606
           || ||| ||   |    |||        ||| | || |      ||  |||| |  || ||
Db     487 GCAGTTAAAGAAGGCGAATTATTATTCGGTGATACTGTAGCAGTATTTGGCGCTGGGCCA 546

Qy     607 ATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCC 666
           || ||  |    || ||| ||||     |||    |  |||  | |||||  |    |
Db     547 ATTGGTTTGTTAACTATTGTTGCAGCAAAAGCTGCTGGTGCAAGTAAAATATTTGTCTTT 606

Qy     667 GAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCT 726
           ||      |||       |  ||  || || || |       | |  | | |    |
Db     607 GA------CTTATCAGAAGAACGTTTAGCGAAAGCTAAAAGTGTCGGTGCGACTCACGTG 660

Qy     727 TCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTC 786
           | ||  |  |  |||   || | || |  | || |   |  |       || | ||  |
Db     661 TATAACTCAGGTAACGTCGATCCAGTACAAACGGTTTATGAACATACTGACAACGGTGTA 720

Qy     787 CATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACG 846
            |||   | || ||     |||||||    |   || ||       || ||||  |  ||
Db     721 GATGTGTCATTTGAAGTTGCTGGTGTAGGTATTACTTTACAACAATCTATTGAAGTAACA 780

Qy     847 GGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCAC 891
           | ||   ||| |  || |||| ||  |  | |  ||  | | |
Db     781 CGTCCACGTGGTACTGCTGTCATCGTATCAATCTTCGGTCATCCC 825


RESULT 3
US-09-347-803-19
; Sequence 19, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Hitz, Bill
;  APPLICANT: Kinney, Tony
;  APPLICANT: Orozco, Buddy
;  TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
;  FILE REFERENCE: BB-1176
;  CURRENT APPLICATION NUMBER: US/09/347,803
;  CURRENT FILING DATE: 1999-07-02
;  EARLIER APPLICATION NUMBER: 60/092,952
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
;   LENGTH: 1610
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-347-803-19

 Query Match             4.3%;  Score 49;  DB 4;  Length 1610;
  Best Local Similarity   48.7%;  Pred. No. 1.4e-05;
  Matches  133;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy     443 TTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCG 502
           | ||  |  |  | ||  |   ||  ||||      | ||||||||||||    |   |
Db     427 TGGCTAATCAGATAGTGCATCCTGCAGACCTATGTTTTAAGCTGCCAGACAACGTGAGCC 486

Qy     503 ACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTA 562
           ||    |||||| ||  ||| ||  |   |||||     ||||||  |  | | ||||
Db     487 TAGAGGAGGGAGCAATGTGTGAACCCTTAAGTGTTGGTGTTCATGCTTGTAGAAGAGCTA 546

Qy     563 GATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCA 622
           |      || |     |  |   || |  |||| ||| || || || ||||  || |
Db     547 ATATTGGACCAGAAACAAATGTGTTGATCATGGGAGCAGGGCCCATAGGACTTGTTACAA 606

Qy     623 TTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCA 682
           | || ||    |  |    |   ||   |||||  ||  |   ||    |
Db     607 TGCTGGCAGCTCGTGCGTTTGGGGCACCCAAAACAGTCATTGTGGATGTTGATGACCATC 666

Qy     683 GAAGACAGTTTGCAAAGGAACTGGGCGCTGAAG 715
           |   |    |||||||    || || || || |
Db     667 GTTTATCTGTTGCAAAATCTCTTGGTGCAGATG 699


RESULT 4
US-08-466-548B-1
; Sequence 1, Application US/08466548B
; Patent No. 6022712
;  GENERAL INFORMATION:
;    APPLICANT:  sarthy, aparna v
;    APPLICANT:  schopp, cynthia w
;    TITLE OF INVENTION:  ENHANCED YEAST EXPRESSION USING
;    TITLE OF INVENTION:  REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
;    TITLE OF INVENTION:  DEHYDROGENASE GENE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LAORTORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,548B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/998,226
;      FILING DATE:  30-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WALSH, ANDREA C
;      REGISTRATION NUMBER:  34988
```

```
;       REFERENCE/DOCKET NUMBER:  5283.us.01
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2774 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  SACCHAROMYCES CEREVISIAE
;       STRAIN:  800B
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  788..1856
US-08-466-548B-1

  Query Match             4.1%;  Score 47.2;  DB 3;  Length 2774;
  Best Local Similarity   72.6%;  Pred. No. 7.2e-05;
  Matches   61;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     220 GGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCT 279
           || |||||||  ||||||  ||| || || |||||   ||  |||||||| || |||  |
Db     986 GGTCATGAATCAAGCGGACAGGTTGTGGAAGTTGGTGATGCCGTCACAAGGGTCAAAGTT 1045

Qy     280 GGTGACAGAGTCGCAGTTGAAGCT 303
           |||||| | || ||  ||||| ||
Db    1046 GGTGACCGTGTTGCTATTGAACCT 1069


RESULT 5
US-07-998-226F-1
; Sequence 1, Application US/07998226F
; Patent No. 6033898
; GENERAL INFORMATION:
;   APPLICANT:  sarthy, aparna v
;   APPLICANT:  schopp, cynthia w
;   TITLE OF INVENTION:  ENHANCED YEAST EXPRESSION USING
;   TITLE OF INVENTION:  REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
;   TITLE OF INVENTION:  DEHYDROGENASE GENE
;   NUMBER OF SEQUENCES:  9
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  ABBOTT LAORTORIES
;     STREET:  ONE ABBOTT PARK ROAD
;     CITY:  ABBOTT PARK
;     STATE:  ILLINOIS
;     COUNTRY:  US
;     ZIP:  60064-3500
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/998,226F
;     FILING DATE:  30-DEC-1992
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  WALSH, ANDREA C
;     REGISTRATION NUMBER:  34988
;     REFERENCE/DOCKET NUMBER:  5283.us.01
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2774 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   ORIGINAL SOURCE:
;     ORGANISM:  SACCHAROMYCES CEREVISIAE
;     STRAIN:  800B
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  788..1856
US-07-998-226F-1

  Query Match             4.1%;  Score 47.2;  DB 3;  Length 2774;
  Best Local Similarity   72.6%;  Pred. No. 7.2e-05;
  Matches   61;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     220 GGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCT 279
           || |||||||  ||||||  ||| || || |||||   ||  |||||||| || |||  |
Db     986 GGTCATGAATCAAGCGGACAGGTTGTGGAAGTTGGTGATGCCGTCACAAGGGTCAAAGTT 1045

Qy     280 GGTGACAGAGTCGCAGTTGAAGCT 303
           |||||| | || ||  ||||| ||
Db    1046 GGTGACCGTGTTGCTATTGAACCT 1069


RESULT 6
PCT-US93-12560-1
; Sequence 1, Application PC/TUS9312560
; GENERAL INFORMATION:
;   APPLICANT:  sarthy, aparna v
;   APPLICANT:  schopp, cynthia w
;   TITLE OF INVENTION:  ENHANCED YEAST EXPRESSION IN USING
;   TITLE OF INVENTION:  REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
;   TITLE OF INVENTION:  DEHYDROGENASE GENE
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  ABBOTT LAORTORIES
;     STREET:  ONE ABBOTT PARK ROAD
;     CITY:  ABBOTT PARK
;     STATE:  ILLINOIS
;     COUNTRY:  US
;     ZIP:  60064-3500
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US93/12560
;     FILING DATE:  22-DEC-1993
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/998,226
;     FILING DATE:  30-DEC-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  WALSH, ANDREA C
;     REGISTRATION NUMBER:  34988
;     REFERENCE/DOCKET NUMBER:  5283.us.01
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2774 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   ORIGINAL SOURCE:
;     ORGANISM:  SACCHAROMYCES CEREVISIAE
;     STRAIN:  800B
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  788..1856
PCT-US93-12560-1

  Query Match             4.1%;  Score 47.2;  DB 5;  Length 2774;
  Best Local Similarity   72.6%;  Pred. No. 7.2e-05;
  Matches   61;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;
```

```
Qy    220 GGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCT 279
          || ||||||| |||||| ||| || || ||||| || ||||||| || ||| |
Db    986 GGTCATGAATCAAGCGGACAGGTTGTGGAAGTTGGTGATGCCGTCACAAGGGTCAAAGTT 1045

Qy    280 GGTGACAGAGTCGCAGTTGAAGCT 303
          |||||| | || || ||||| ||
Db   1046 GGTGACCGTGTTGCTATTGAACCT 1069


RESULT 7
US-08-174-467-18
; Sequence 18, Application US/08174467
; Patent No. 5451514
;  GENERAL INFORMATION:
;    APPLICANT:  BOUDET, ALAIN M.
;    APPLICANT:  INZE, DIRK G.
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/174,467
;      FILING DATE:  28-DEC-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/874,166
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  95563/PS36321/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-174-467-18

  Query Match             3.9%;  Score 45;  DB 1;  Length 1393;
  Best Local Similarity   67.7%;  Pred. No. 0.00024;
  Matches   63;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy    208 CCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACA 267
          |||||   |||||||||||||| |    |||   ||||| ||||| ||  | | |||  ||
Db    272 CCTCTGGTTCCTGGACATGAAGTGGTGGGAGAAGTGGTGGAGGTAGGACCAGATGTGTCA 331

Qy    268 AGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
          |   | |||  ||| |||| ||| | ||||| |
Db    332 AAATTCAAAGTTGGGGACACAGTTGGAGTTGGA 364


RESULT 8
US-08-452-071-18
; Sequence 18, Application US/08452071
; Patent No. 6066780
;  GENERAL INFORMATION:
;    APPLICANT:  BOUDET, ALAIN M.
;    APPLICANT:  INZE, DIRK G.
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,071
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/874,166
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  95563/PS36321/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-452-071-18

  Query Match             3.9%;  Score 45;  DB 3;  Length 1393;
  Best Local Similarity   67.7%;  Pred. No. 0.00024;
  Matches   63;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy    208 CCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACA 267
          |||||   |||||||||||||| |    |||   ||||| ||||| ||  | | |||  ||
Db    272 CCTCTGGTTCCTGGACATGAAGTGGTGGGAGAAGTGGTGGAGGTAGGACCAGATGTGTCA 331

Qy    268 AGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
          |   | |||  ||| |||| ||| | ||||| |
Db    332 AAATTCAAAGTTGGGGACACAGTTGGAGTTGGA 364


RESULT 9
US-08-174-467-17
; Sequence 17, Application US/08174467
; Patent No. 5451514
;  GENERAL INFORMATION:
;    APPLICANT:  BOUDET, ALAIN M.
;    APPLICANT:  INZE, DIRK G.
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/174,467
;       FILING DATE:  28-DEC-1993
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/874,166
;       FILING DATE:  27-APR-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOKULIS, PAUL N.
;       REGISTRATION NUMBER:  16,773
;       REFERENCE/DOCKET NUMBER:  95563/PS36321/US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-861-3000
;       TELEFAX:  202-822-0944
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1419 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-174-467-17

  Query Match             3.8%;  Score 43.4;  DB 1;  Length 1419;
  Best Local Similarity   66.7%;  Pred. No. 0.0008;
  Matches   62;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy     208 CCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACA 267
           |||||  |||||||||||||| |    ||||  || || ||||| ||  | | |||  ||
Db     281 CCTCTGGTTCCTGGACATGAAGTGGTGGGAAAAGTAGTGGAGGTAGGAGCAGATGTGTCA 340

Qy     268 AGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
           |   | |||   || |||| ||| | ||||| |
Db     341 AAATTCAAAGTGGGGGACACAGTTGGAGTTGGA 373


RESULT 10
US-08-452-071-17
; Sequence 17, Application US/08452071
; Patent No. 6066780
;  GENERAL INFORMATION:
;    APPLICANT:  BOUDET, ALAIN M.
;    APPLICANT:  INZE, DIRK G.
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,071
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/874,166
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  95563/PS36321/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1419 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-452-071-17

  Query Match             3.8%;  Score 43.4;  DB 3;  Length 1419;
  Best Local Similarity   66.7%;  Pred. No. 0.0008;
  Matches   62;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy     208 CCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACA 267
           |||||  |||||||||||||| |    ||||  || || ||||| ||  | | |||  ||
Db     281 CCTCTGGTTCCTGGACATGAAGTGGTGGGAAAAGTAGTGGAGGTAGGAGCAGATGTGTCA 340

Qy     268 AGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
           |   | |||   || |||| ||| | ||||| |
Db     341 AAATTCAAAGTGGGGGACACAGTTGGAGTTGGA 373


RESULT 11
US-09-347-803-15
; Sequence 15, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Hitz, Bill
;  APPLICANT: Kinney, Tony
;  APPLICANT: Orozco, Buddy
;  TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
;  FILE REFERENCE: BB-1176
;  CURRENT APPLICATION NUMBER: US/09/347,803
;  CURRENT FILING DATE: 1999-07-02
;  EARLIER APPLICATION NUMBER: 60/092,952
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1154
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-347-803-15

  Query Match             3.7%;  Score 41.8;  DB 4;  Length 1154;
  Best Local Similarity   50.2%;  Pred. No. 0.0023;
  Matches  103;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     435 CGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTC 494
           ||||    | ||  |  | || ||  ||    |  ||||       ||||||| || |||
Db     501 CGGCTCGCTGGCGAACCAGGTGGTGCACCCGGCCGACCTGTGCTTCAAGCTCCCCGACGG 560

Qy     495 GATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGA 554
           | |   |   ||    || ||  ||   |||||  |    || | |  ||| ||
Db     561 GGTGAGCCTGGAGGAGGGCGCCATGTGCGAGCCGCTGAGCGTGGGCGTGCACGCGTGCCG 620

Qy     555 ACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCT 614
           |||||  |  |   ||| |    |  |  || ||  | || |  || || ||||||||
Db     621 CCGCGCGGGGGTGGGGCCCGAGACGGGCGTGCTCGTGGTGGGCGCCGGCCCCATCGGCCT 680

Qy     615 TGCCACCATTCTTGCTCTGCAAGGC 639
           |   |  | || ||   || || |
Db     681 GGTGTCGCTGCTAGCGGCGCGAGCC 705


RESULT 12
US-08-174-467-19
; Sequence 19, Application US/08174467
```

```
; Patent No. 5451514
;  GENERAL INFORMATION:
;    APPLICANT:  BOUDET, ALAIN M.
;    APPLICANT:  INZE, DIRK G.
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/174,467
;      FILING DATE:  28-DEC-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/874,166
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  95563/PS36321/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1283 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-174-467-19

  Query Match             3.6%;  Score 40.6;  DB 1;  Length 1283;
  Best Local Similarity   64.2%;  Pred. No. 0.006;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     207 TCCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCAC 266
           |||| |    ||||| |||||| |     ||   ||| || ||||| || || | ||| ||
Db     216 TCCTATGGTCCCTGGCCATGAAGTGGTTGGTGAGGTTGTTGAGGTGGGATCAGATGTGAC 275

Qy     267 AAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAG 301
           |||  | |||  ||| ||    ||||  |||| ||
Db     276 AAGGTTCAAAGTTGGAGATGTTGTCGGTGTTGGAG 310


RESULT 13
US-08-452-071-19
; Sequence 19, Application US/08452071
; Patent No. 6066780
;  GENERAL INFORMATION:
;    APPLICANT:  BOUDET, ALAIN M.
;    APPLICANT:  INZE, DIRK G.
;    APPLICANT:  SCHUCH, WOLFGANG W.
;    TITLE OF INVENTION:  MODIFICATION OF LIGNIN SYNTHESIS IN
;    TITLE OF INVENTION:  PLANTS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,071
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/874,166
;      FILING DATE:  27-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  95563/PS36321/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1283 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-452-071-19

  Query Match             3.6%;  Score 40.6;  DB 3;  Length 1283;
  Best Local Similarity   64.2%;  Pred. No. 0.006;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     207 TCCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCAC 266
           |||| |    ||||| |||||| |     ||   ||| || ||||| || || | ||| ||
Db     216 TCCTATGGTCCCTGGCCATGAAGTGGTTGGTGAGGTTGTTGAGGTGGGATCAGATGTGAC 275

Qy     267 AAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAG 301
           |||  | |||  ||| ||    ||||  |||| ||
Db     276 AAGGTTCAAAGTTGGAGATGTTGTCGGTGTTGGAG 310


RESULT 14
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,468
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
```

```
;       REGISTRATION NUMBER:  31,145
;       REFERENCE/DOCKET NUMBER:  31,255-02
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (201)831-3241
;       TELEFAX:  (201)831-3305
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  30001 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-125-468-1

  Query Match             3.5%;  Score 39.8;  DB 1;  Length 30001;
  Best Local Similarity   55.4%;  Pred. No. 0.072;
  Matches   77;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     466 GAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAG 525
           | |||||||   | | |||| |||| |  | ||  || ||   | | ||| | |||||||
Db   14764 GCGGACCACGGCGGCCAGCTTCCAGCCGAGGTTGACGGCGTCCTGGAAGCCCAGGTTGAG 14705

Qy     526 CCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCC 585
           ||  |  | |  | | |||| |  | ||  ||||       |   |     ||  |||||
Db   14704 CCCCTGGCCGCCGGCCGGCAGGTGGATGTGCGCGGCGTCGCCGGCCAGCAGCACCCGGCC 14645

Qy     586 CTGGTTCTTGGAGGAGGTC 604
           | |    | ||||  || |
Db   14644 CAGCCGGTAGGAGTCGGCC 14626


RESULT 15
US-08-474-933-1/c
; Sequence 1, Application US/08474933
; Patent No. 5866410
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,933
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,468
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255-02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-933-1

  Query Match             3.5%;  Score 39.8;  DB 2;  Length 30001;
  Best Local Similarity   55.4%;  Pred. No. 0.072;
  Matches   77;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     466 GAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAG 525
           | |||||||   | | |||| |||| |  | ||  || ||   | | ||| | |||||||
Db   14764 GCGGACCACGGCGGCCAGCTTCCAGCCGAGGTTGACGGCGTCCTGGAAGCCCAGGTTGAG 14705

Qy     526 CCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCC 585
           ||  |  | |  | | |||| |  | ||  ||||       |   |     ||  |||||
Db   14704 CCCCTGGCCGCCGGCCGGCAGGTGGATGTGCGCGGCGTCGCCGGCCAGCAGCACCCGGCC 14645

Qy     586 CTGGTTCTTGGAGGAGGTC 604
           | |    | ||||  || |
Db   14644 CAGCCGGTAGGAGTCGGCC 14626


Search completed:  May  2, 2003, 09:34:38
Job time :  153 secs
```

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 09:10:26 ; Search time 171 Seconds
(without alignments)
7898.204 Million cell updates/sec

Title: US-10-020-674-1
Perfect score: 1143
Sequence: 1 atgaaaggtttactttatta.....cgccgaacgaggtttcctaa 1143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1143	100.0	1143	9	US-10-020-674-1	Sequence 1, Appli
	2	515.8	45.1	530	9	US-10-020-674-8	Sequence 8, Appli
	3	402.4	35.2	706	9	US-10-020-674-12	Sequence 12, Appl
	4	328	28.7	523	9	US-10-020-674-15	Sequence 15, Appl
	5	317.4	27.8	1158	9	US-10-147-003-1	Sequence 1, Appli
	6	193.2	16.9	745	9	US-10-147-003-7	Sequence 7, Appli
	7	103	9.0	107	9	US-10-020-674-11	Sequence 11, Appl
	8	83	7.3	1041	9	US-09-891-641-11	Sequence 11, Appl
	9	63.8	5.6	1092	10	US-09-815-242-7881	Sequence 7881, Ap
	10	52.2	4.6	2410	10	US-09-925-300-386	Sequence 386, App
	11	52.2	4.6	2519	10	US-09-969-708-281	Sequence 281, App
	12	52.2	4.6	2519	10	US-09-880-107-2189	Sequence 2189, Ap
	13	45.8	4.0	1014	10	US-09-974-300-551	Sequence 551, App
	14	45	3.9	513	9	US-10-025-380-153	Sequence 153, App
	15	45	3.9	513	10	US-09-922-217-153	Sequence 153, App
	16	45	3.9	513	10	US-09-833-263-153	Sequence 153, App
c	17	45	3.9	14333	7	US-08-781-986A-57	Sequence 57, Appl
	18	42	3.7	2277	10	US-09-917-800A-1513	Sequence 1513, Ap
	19	41.8	3.7	915	10	US-09-815-242-4651	Sequence 4651, Ap
	20	41.8	3.7	1044	10	US-09-815-242-8368	Sequence 8368, Ap
	21	39.8	3.5	1059	10	US-09-815-242-7877	Sequence 7877, Ap
	22	39.2	3.4	1074	9	US-09-938-842A-1405	Sequence 1405, Ap
	23	39	3.4	438	10	US-09-974-300-4985	Sequence 4985, Ap
	24	39	3.4	1395	10	US-09-947-027-8	Sequence 8, Appli
	25	39	3.4	1395	12	US-10-091-009-8	Sequence 8, Appli
	26	39	3.4	1400	9	US-09-908-744-68	Sequence 68, Appl
	27	38	3.3	451	10	US-09-770-444-670	Sequence 670, App
	28	37.8	3.3	453	10	US-09-974-300-4984	Sequence 4984, Ap
	29	37.4	3.3	990	9	US-09-938-842A-824	Sequence 824, App
	30	37	3.2	1049	9	US-10-123-155-358	Sequence 358, App
	31	35.8	3.1	1573	10	US-09-971-361-8	Sequence 8, Appli
	32	35.6	3.1	395	9	US-09-803-719-859	Sequence 859, App
	33	35.6	3.1	400	9	US-09-803-719-819	Sequence 819, App
c	34	35.6	3.1	992	10	US-09-815-242-2647	Sequence 2647, Ap
	35	35.6	3.1	1560	7	US-08-781-986A-90	Sequence 90, Appl
c	36	35.2	3.1	451	9	US-09-918-995-2486	Sequence 2486, Ap
	37	35.2	3.1	1015	10	US-09-974-300-552	Sequence 552, App
c	38	35.2	3.1	2061	12	US-10-096-960-1	Sequence 1, Appli
	39	34.6	3.0	434	9	US-10-079-623-245	Sequence 245, App
	40	34.6	3.0	451	10	US-09-960-352-8378	Sequence 8378, Ap
	41	34.4	3.0	762	9	US-09-738-626-2710	Sequence 2710, Ap
	42	34	3.0	585	9	US-09-991-936-1554	Sequence 1554, Ap
	43	34	3.0	644	10	US-09-833-381-1031	Sequence 1031, Ap
	44	34	3.0	1011	10	US-09-815-242-8332	Sequence 8332, Ap
	45	34	3.0	1017	10	US-09-815-242-4441	Sequence 4441, Ap

ALIGNMENTS

RESULT 1
US-10-020-674-1
; Sequence 1, Application US/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Onodera, Keiko
; APPLICANT: Tani, Yoshiki
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
; FILE REFERENCE: 06501-092001
; CURRENT APPLICATION NUMBER: US/10/020,674
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: JP 2000-333363
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Pichia angusta
US-10-020-674-1

Query Match 100.0%; Score 1143; DB 9; Length 1143;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60

Qy      61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120

Qy     121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180

Qy     181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
```

```
Qy    241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300

Qy    301 GCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGG 360

Qy    361 CTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGT 420

Qy    421 TTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTC 480

Qy    481 AAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCC 540

Qy    541 TGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGA 600

Qy    601 GGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTG 660

Qy    661 TGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTC 720

Qy    721 GATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAG 780

Qy    781 GGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTC 840

Qy    841 GCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGA 900

Qy    901 TTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACC 960

Qy    961 GTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAA 1020

Qy   1021 GCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAA 1080

Qy   1081 CAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACGAGGTTTCC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACGAGGTTTCC 1140

Qy   1141 TAA 1143
          |||
Db   1141 TAA 1143
```

RESULT 2
US-10-020-674-8
; Sequence 8, Application US/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Onodera, Keiko
; APPLICANT: Tani, Yoshiki
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
; FILE REFERENCE: 06501-092001
; CURRENT APPLICATION NUMBER: US/10/020,674
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: JP 2000-333363
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Pichia angusta
US-10-020-674-8

Query Match 45.1%; Score 515.8; DB 9; Length 530;
Best Local Similarity 98.7%; Pred. No. 4.6e-164;
Matches 520; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy    274 AAACCTGGTGACAGAGTCGCAGTTGAAGCTACGTCGCATTGCTCCGACAGATCGCGCTAC 333
          || || ||||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCCGGGTGATCGTGTCGCAGTTGAAGCTACGTCGCATTGCTCCGACAGATCGCGCTAC 60

Qy    334 AAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAAC 393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAAC 120

Qy    394 TGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTAC 453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTAC 180

Qy    454 GTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGA 513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGA 240

Qy    514 GCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCT 573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCT 300

Qy    574 GGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTG 633
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTG 360

Qy    634 CAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTT 693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTT 420

Qy    694 GCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTT 753
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTT 480

Qy    754 CTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCGA 800
          ||||||||||||||||||||||||||||||||||| || ||||||||
Db    481 CTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCACGCCGCCTTCGA 527
```

RESULT 3
US-10-020-674-12
; Sequence 12, Application US/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Onodera, Keiko
; APPLICANT: Tani, Yoshiki
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
; FILE REFERENCE: 06501-092001
; CURRENT APPLICATION NUMBER: US/10/020,674
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: JP 2000-333363
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 706

```
;   TYPE: DNA
;   ORGANISM: Pichia angusta
US-10-020-674-12

  Query Match             35.2%;  Score 402.4;  DB 9;  Length 706;
  Best Local Similarity   99.8%;  Pred. No. 1.4e-125;
  Matches  403;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    740 ACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCG 799
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 ATGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGCAGCCTTCG 62

Qy    800 ACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAA 859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 ACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACCTTCTGGAA 122

Qy    860 TCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCTGA 919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 TCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAATGTCTCTGA 182

Qy    920 CTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGAAG 979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 CTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTTCCAGGAAG 242

Qy    980 TTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGATGATTACAG 1039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 TTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGATGATTACAG 302

Qy   1040 GCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGAGCACAAGG 1099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 GCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGAGCACAAGG 362

Qy   1100 AGAACAATGTCAAGATCCTGGTGACGCCGAACGAGGTTTCCTAA 1143
          ||||||||||||||||||||||||||||||||||||||||||||
Db    363 AGAACAATGTCAAGATCCTGGTGACGCCGAACGAGGTTTCCTAA 406

:

RESULT 4
US-10-020-674-15
; Sequence 15, Application US/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
;  APPLICANT: Yamamoto, Hiroaki
;  APPLICANT:  Onodera, Keiko
;  APPLICANT:  Tani, Yoshiki
;  TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
;  FILE REFERENCE: 06501-092001
;  CURRENT APPLICATION NUMBER: US/10/020,674
;  CURRENT FILING DATE:  2002-04-19
;  PRIOR APPLICATION NUMBER: JP 2000-333363
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 523
;   TYPE: DNA
;   ORGANISM: Pichia angusta
US-10-020-674-15

  Query Match             28.7%;  Score 328;  DB 9;  Length 523;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-100;
  Matches  328;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 251

Qy     61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 311

Qy    121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 371

Qy    181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 431

Qy    241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 491

Qy    301 GCTACGTCGCATTGCTCCGACAGATCGC 328
          ||||||||||||||||||||||||||||
Db    492 GCTACGTCGCATTGCTCCGACAGATCGC 519


RESULT 5
US-10-147-003-1
; Sequence 1, Application US/10147003
; Publication No. US20030032153A1
; GENERAL INFORMATION:
;  APPLICANT: Yamamoto, Hiroaki
;  APPLICANT:  Kimoto, No. US20030032153Alihiro
;  TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE,
;  TITLE OF INVENTION:  METHODS FOR PRODUCING SAME, AND METHODS FOR PRODUCING
;  TITLE OF INVENTION:  OPTICALLY ACTIVE ALCOHOL USING THE DEHYDROGENASE
;  FILE REFERENCE: 06501-109001
;  CURRENT APPLICATION NUMBER: US/10/147,003
;  CURRENT FILING DATE:  2002-05-16
;  PRIOR APPLICATION NUMBER: JP 2001-159647
;  PRIOR FILING DATE: 2001-05-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1158
;   TYPE: DNA
;   ORGANISM: Kluyveromyces lactis
US-10-147-003-1

  Query Match             27.8%;  Score 317.4;  DB 9;  Length 1158;
  Best Local Similarity   56.3%;  Pred. No. 1.2e-96;
  Matches  638;  Conservative    0;  Mismatches  486;  Indels    9;  Gaps    2;

Qy      1 ATGAAAGGTTTACTTTATTACGGTACAAACGATATTCGCTACTCCGAAACGGTTCCTGAA 60
          |||   |  |||  ||||| ||| | | | || ||  | ||| |  |     |   |||
Db      1 ATGCGTGCATTAGCTTATTTCGGAAAACAAGACATCAGATACACAAAGGATTTGGAGGAA 60

Qy     61 CCGGAGATCAAGAATCCCAACGATGTCAAGATCAAAGTCAGCTATTGTGGAATCTGTGGC 120
          || |  ||| | |       | |   |  | ||| |||||   |  ||||| ||||| ||
Db     61 CCTGTCATCGAAACAGATGATGGAATTGAAATCGAAGTCTCATGGTGTGGTATCTGCGGT 120

Qy    121 ACGGACTTGAAAGAATTCACATATTCTGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACC 180
          |  || ||  | |||| |  | ||      ||||||  |||| || ||  || | ||||
Db    121 AGTGATTTACACGAATACCTAGAT------GGTCCTATTTTCTTTCCAGAAGATGGCAAG 174

Qy    181 AAGGACAAGATTTCGGGATACGAACTTCCTCTCTGTCCTGGACATGAATTTAGCGGAACG 240
            || |  ||   ||    | | | ||||    |   || |||||  |   ||| |
Db    175 GTCCACGATGTTAGTGGTCTTGGATTGCCTCAAGCTATGGGTCATGAGATGTCCGGTATC 234

Qy    241 GTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAA 300
          ||     | |||||  |    || || |   | ||  |||||||    || |  || |||
Db    235 GTATCAAAAGTTGGACCAAAAGTAACCAACATCAAGGCTGGTGATCATGTTGTTGTAGAA 294

Qy    301 GCTACGTCGCATTGCTCCGACAGATCGCGCTACAAGGACACGGTCGCCCAAGACCTTGGG 360
          || ||        ||    |    |      || | |          |   ||
Db    295 GCCACCGGTACATGTCTTGATCATTACACTTGGCCTAACGCTGCACATGCTAAGGATGCT 354

Qy    361 CTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGT 420
          |||   || || || || || |      ||||| |||||     || | |||    |||
Db    355 GAATGTGCTGCTTGTCAAAGAGGGTTCTACAACTGTTGTGCCCATTTGGGTTTCATGGGT 414

Qy    421 TTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTC 480
          | || | |  ||| || |||||||| ||  | ||||||    ||||  | ||| | ||
Db    415 CTAGGAGTTCACAGTGGTGGTTTTGCTGAAAAAGTCGTTGTTAGTGAAAAGCACGTTGTT 474
```

```
Qy    481 AAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCC 540
Db    475 AAGATTCCAAACACTTTACCATTGGACGTTGCAGCCTTAGTCGAACCAATCTCCGTCTCA 534

Qy    541 TGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGA 600
Db    535 TGGCATGCTGTTAGAATCTCCAAGTTACAGAAGGGCCAATCCGCCTTAGTTCTTGGCGCA 594

Qy    601 GGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTG 660
Db    595 GGTCCAATTGGATTAGCCACCATTTTAGCTTTGCAAGGTCATGGTGCCTCCAAGATTGTA 654

Qy    661 TGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTC 720
Db    655 GTCTCTGAACCAGCTGAAATCAGAAGAAATCAAGCTGCCAAGCTTGGTGTCGAAACCTTC 714

Qy    721 GATCCTTCTACATGT---GACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAAC 777
Db    715 GATCCTTCTGAACATAAAGAAGATGCCGTTAATATTTTGAAGAAATTGGCTCCAGGTGGT 774

Qy    778 GAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATT 837
Db    775 GAAGGTTTCGATTTTGCCTACGATTGTTCTGGTGTTAAACCAACTTTTGATACAGGTGTC 834

Qy    838 GTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATT 897
Db    835 CATGCTACTACCTTCAGAGGTATGTATGTTAATATTGCAATCTGGGGCCACAAACCCATC 894

Qy    898 GGATTCATGCCAATGTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTAC 957
Db    895 GATTTCAAACCTATGGACGTCACTCTGCAAGAGAAATTCGTCACCGGTTCCATGTGTTAC 954

Qy    958 ACCGTCAAGGACTTCCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGAC 1017
Db    955 ACCATTAAAGATTTTGAAGATGTGGTCCAAGCCTTAGGCAATGGCAGCATTGCTATTGAT 1014

Qy   1018 AAAGCGCGCAAGATGATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTT 1077
Db   1015 AAAGCTAGACATTTGATTACGGGTAGACAAAAAATCGAAGATGGGTTCACTAAAGGGTTC 1074

Qy   1078 AAACAGCTGATCGAGCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAA 1130
Db   1075 GACGAATTAATGAACCATAAAGAGAAAAACATCAAGATATTGTTGACTCCTAA 1127
```

RESULT 6
US-10-147-003-7
; Sequence 7, Application US/10147003
; Publication No. US20030032153A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20030032153A1ihiro
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE,
; TITLE OF INVENTION: METHODS FOR PRODUCING SAME, AND METHODS FOR PRODUCING
; TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING THE DEHYDROGENASE
; FILE REFERENCE: 06501-109001
; CURRENT APPLICATION NUMBER: US/10/147,003
; CURRENT FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: JP 2001-159647
; PRIOR FILING DATE: 2001-05-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Kluyveromyces aestuarii
US-10-147-003-7

Query Match 16.9%; Score 193.2; DB 9; Length 745;
Best Local Similarity 55.4%; Pred. No. 1e-54;
Matches 397; Conservative 0; Mismatches 313; Indels 6; Gaps 1;

```
Qy     27 AAACGATATTCGCTACTCCGAAACGGTTCCTGAACCGGAGATCAAGAATCCCAACGATGT 86
Db      2 ACAAGACATCAGATACACAAAGGATTTGGAGGAACCTGTCATCGAAACAGATGATGGAAT 61

Qy     87 CAAGATCAAAGTCAGCTATTGTGGAATCTGTGGCACGGACTTGAAAGAATTCACATATTC 146
Db     62 TGAAATCGAAGTCTCATGGTGTGGTATCTGCGGTAGTGATTTACACGAATACCTAGAT-- 119

Qy    147 TGGAGGTCCTGTTTTTTTCCCTAAACAAGGCACCAAGGACAAGATTTCGGGATACGAACT 206
Db    120 ----GGTCCTATTTTCTTTCCAGAAGATGGCAAGGTCCACGATGTTAGTGGTCTTGGATT 175

Qy    207 TCCTCTCTGTCCTGGACATGAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCAC 266
Db    176 GCCTCAAGCTATGGGTCATGAGATGTCCGGTATCGTATCAAAAGTTGGACCAAAAGTAAC 235

Qy    267 AAGTGTGAAACCTGGTGACAGAGTCGCAGTTGAAGCTACGTCGCATTGCTCCGACAGATC 326
Db    236 CAACATCAAGGCTGGTGATCATGTTGTTGTAGAAGCCACCGGTACATGTCTTGATCATTA 295

Qy    327 GCGCTACAAGGACACGGTCGCCCAAGACCTTGGGCTCTGTATGGCCTGCCAGAGCGGATC 386
Db    296 CACTTGGCCTAACGCTGCACATGCTAAGGATGCTGAATGTGCTGCTTGTCAAAGAGGGTT 355

Qy    387 TCCGAACTGCTGTGCGTCGCTGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGC 446
Db    356 CTACAACTGTTGTGCCCATTTGGGTTTCATGGGTCTAGGAGTTCACAGTGGTGGTTTTGC 415

Qy    447 CGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGA 506
Db    416 TGAAAAAGTCGTTGTTAGTGAAAAGCACGTTGTTAAGATTCCAAACACTTTACCATTGGA 475

Qy    507 TATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATT 566
Db    476 CGTTGCAGCCTTAGTCGAACCAATCTCCGTCTCATGGCATGCTGTTAGAATCTCCAAGTT 535

Qy    567 CCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCT 626
Db    536 ACAGAAGGGCCAATCCGCCTTAGTTCTTGGCGCAGGTCCAATTGGATTAGCCACCATTTT 595

Qy    627 TGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAG 686
Db    596 AGCTTTGCAAGGTCATGGTGCCTCCAAGATTGTAGTCTCTGAACCAGCTGAAATCAGAAG 655

Qy    687 ACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTACATGTGACGACG 742
Db    656 AAATCAAGCTGCCAAGCTTGGTGTCGAAACCTTCGATCCTTCTGAACATAAAGAAG 711
```

RESULT 7
US-10-020-674-11
; Sequence 11, Application US/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Onodera, Keiko
; APPLICANT: Tani, Yoshiki
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
; FILE REFERENCE: 06501-092001
; CURRENT APPLICATION NUMBER: US/10/020,674
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: JP 2000-333363
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 107
; TYPE: DNA
; ORGANISM: Pichia angusta
US-10-020-674-11

Query Match 9.0%; Score 103; DB 9; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;

```
  Matches  103;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     226 GAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGAC 285
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAATTTAGCGGAACGGTGGTCGAGGTTGGCTCTGGTGTCACAAGTGTGAAACCTGGTGAC 60

Qy     286 AGAGTCGCAGTTGAAGCTACGTCGCATTGCTCCGACAGATCGC 328
           |||||||||||||||||||||||||||||||||||||||||||
Db      61 AGAGTCGCAGTTGAAGCTACGTCGCATTGCTCCGACAGATCGC 103


RESULT 8
US-09-891-641-11
; Sequence 11, Application US/09891641
; Patent No. US20020155612A1
; GENERAL INFORMATION:
;  APPLICANT: Ye, Rick
;  APPLICANT:  Bedzyk, Laura
;  APPLICANT:  Wang, Tao
;  TITLE OF INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN BACILLUS SPECIES
;  FILE REFERENCE: CL1686 US NA
;  CURRENT APPLICATION NUMBER: US/09/891,641
;  CURRENT FILING DATE:  2001-06-26
;  NUMBER OF SEQ ID NOS: 81
;  SOFTWARE: Microsoft 97
; SEQ ID NO 11
;   LENGTH: 1041
;   TYPE: DNA
;   ORGANISM: Bacillus subtilis
US-09-891-641-11

  Query Match             7.3%;  Score 83;  DB 9;  Length 1041;
  Best Local Similarity   49.9%;  Pred. No. 2.7e-17;
  Matches  240;  Conservative    0;  Mismatches  235;  Indels    6;  Gaps    1;

Qy     407 TGAGCTTCTGCGGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTG 466
           || | |||  ||| || |  || |  ||||||||||  | || ||||||  |   | ||
Db     338 TGGGATTCCTCGGCTTAGCCGGCGGAGGCGGCGGTTTCTCTGAATACGTCTCTGTGGATG 397

Qy     467 AGGACCACATGGTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGC 526
           | || |   || |||| || || ||     |  |  | ||    || || || ||||| |
Db     398 AAGAGCTTTTGTTCAAACTTCCTGATGAATTATCATATGAACAAGGCGCGCTCGTTGAAC 457

Qy     527 CTATTTCTGTTGCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCC 586
           ||  | | |||||    | |||||          | | || |  | ||  | |  ||
Db     458 CTTCTGCAGTTGCTCTATACGCTGTCCGCTCAAGCAAACTCAAAGCAGGCGACAAAGCGG 517

Qy     587 TGGTTCTTGGAGGAGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATG 646
           ||  | ||  | || || ||||| |||     ||||  || |||·| |        ||
Db     518 CTGTATTCGGCTGCGGCCCGATCGGACTTCTTGTCATTGAAGCGCTGAAGGCTGCCGGTG 577

Qy     647 CGGGCAAAATTGTGTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGG 706
           |    | |||     |  ||||  |       |    ||   || |||| || |
Db     578 CAACTGATATTTACGCTGTTGAGCTTTCTCCTGAACGCCAGCAAAAAGCTGAGGAGCTTG 637

Qy     707 GCGCTGAAGTGTTCGATCCTTCTACATGTGACGACGCAAATGCTGTTCTCAAGGCTATGG 766
           ||||     |  | ||||| |||| |   |||||  | |   ||||       ||  |   |
Db     638 GCGCGATCATCGTTGATCCGTCTAAAACAGACGATGTAGTCGCTG------AGATTGCAG 691

Qy     767 TGCCGGAGAACGAGGGATTCCATGCAGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCA 826
           |      |  ||  |  | ||| |||||   | |||||| ||    |   ||
Db     692 AACGTACAGGAGGCGGTGTTGACGTAGCATTCGAAGTCACTGGTGTCCCAGTGGTGTTAC 751

Qy     827 CCACCTCAATTGTCGCCACGGGACCTTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAG 886
           | ||     ||||   | | | ||     ||||||  ||   | ||||||
Db     752 GACAAGCCATCCAGTCCACTACAATTGCCGGTGAAACCGTCATCGTCAGCATTTGGGAAA 811

Qy     887 A 887
           |
Db     812 A 812


RESULT 9
```

```
US-09-815-242-7881
; Sequence 7881, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7881
;   LENGTH: 1092
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1092)
US-09-815-242-7881

  Query Match             5.6%;  Score 63.8;  DB 10;  Length 1092;
  Best Local Similarity   46.5%;  Pred. No. 8.7e-11;
  Matches  326;  Conservative    0;  Mismatches  357;  Indels   18;  Gaps    3;

Qy     432 CAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATGGTCAAGCTGCCAGA 491
           || |||||  || |||||  ||||||          |   |||   |   |  |||||  |
Db     405 CAACGGCGCCTTCGCCGAATACGTCAACGTACCGGCCAACCTGCTCTACGCCCTGCCGGC 464

Qy     492 CTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGCCTGGCATGCTGT 551
           |    |||||  |||    || || ||| | |||||  |  | || | |  ||| || ||
Db     465 CGGATTTCCCAGCGAGGCCGGCGCGCTGATCGAGCCGCTGGCGGTGGGCATGCACGCGGT 524

Qy     552 TGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGGAGGTCCTATCGG 611
           |    ||  |   || ||  || |||| |   ||||  | || |  ||  | |||||
Db     525 GAAGAAGGCCGGCAGCCTGCTCGGGCAGAACGTGGTGGTGGTCGGCGCCGGCACCATCGG 584

Qy     612 CCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGTGTGTTCCGAGCC 671
           |||  ||||||  |      ||  | |      ||||  |  |  |      |||
Db     585 CCTGAGCACCATCATGTGCGCGCGCGCCGCCGGCGCGGCACAGGTGATCGCCCTGGAGAT 644

Qy     672 GGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTTCGATCCTTCTAC 731
           | ||| |   | |          ||  ||| | |||||      | |        | |
Db     645 GTCCTCGGCGCGCAAGGCCAAGGCCCTGGAGGTCGGCGCCAGCCAGGTGCTCGACCCCTC 704

Qy     732 ATGTGACGACGCAAATGCTGTTCTCAAGGCTATGGTGCCGGAGAACGAGGGATTCCATGC 791
           |  ||||||    | |  ||  ||    | | |          ||   | | |
Db     705 GCGGTGCGACGCCCTAGGCGAAATCCGCGC---CCTCACCGGCGGGCTTGGCGCCGACGT 761

Qy     792 AGCCTTCGACTGCTCTGGTGTTCCTCAGACATTCACCACCTCAATTGTCGCCACGGGACC 851
           ||||||  |||   ||       || ||   ||       | || |  ||||   |
Db     762 CAGCTTCGAATGCATCGGCAACAAGCACACCGCCAAGCTGGCCATCGATGCCATCCGCAA 821
```

```
Qy    852 TTCTGGAATCGCCGTCAATGTGGCCGTTTGGGGAGACCACCCAATTGGATTCATGCCAAT 911
Db    822 GGCCGGCAAATGCGTGCTTGTGGGAATCT---TCGAGGAGCCCAGTGAATTCAATTTCTT 878

Qy    912 GTCTCTGACTTACCAGGAGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTT 971
Db    879 CGAGCTGGTTTCCACCGAGAAGCAGTTGCTCGGGGCCCTCGCCTACAACGGCGAGTTCGC 938

Qy    972 CCAGGAAGTTGTCAAGGCCTTGGAAGATGGTCTCATATCTTTGGACAAAGCGCGCAAGAT 1031
Db    939 CGACGTGATCGCCTTCATAGCCGACGGCCGCCT-----------CGACATCGCACCGCT 986

Qy   1032 GATTACAGGCAAAGTCCACCTAAAGGACGGAGTCGAGAAGGGCTTTAAACAGCTGATCGA 1091
Db    987 GGTGACCGGTCGCATCGGCCTGGAAGAAATCGTCGAGCGCGGTTTCGAAGAGTTGGTGAA 1046

Qy   1092 GCACAAGGAGAACAATGTCAAGATCCTGGTGACGCCGAACG 1132
Db   1047 CAACAAGGAGCACAACGTGAAGATCATCGTCTCGCCGGGCG 1087
```

RESULT 10
US-09-925-300-386
; Sequence 386, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 386
; LENGTH: 2410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2167)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-386

Query Match 4.6%; Score 52.2; DB 10; Length 2410;
Best Local Similarity 51.0%; Pred. No. 1.2e-06;
Matches 123; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy    480 CAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGC 539
Db    421 CAAGCTTCCTGACAATGTCACCTTTGAGGAAGGCGCCCTGATCGAGCCACTTTCTGTGGG 480

Qy    540 CTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGG 599
Db    481 GATCCATGCCTGCAGGAGAGGCGGAGTTACCCTGGGACACAAGGTCCTTGTGTGTGGAGC 540

Qy    600 AGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGT 659
Db    541 TGGGCCAATCGGGATGGTCACTTTGCTCGTGGCCAAAGCAATGGGAGCAGCTCAAGTAGT 600

Qy    660 GTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTT 719
Db    601 GGTGACTGATCTGTCTGCTACCCGATTGTCCAAAGCCAAGGAGATTGGGGCTGATTTAGT 660

Qy    720 C 720
Db    661 C 661
```

RESULT 11
US-09-969-708-281
; Sequence 281, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 281
; LENGTH: 2519
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-281

Query Match 4.6%; Score 52.2; DB 10; Length 2519;
Best Local Similarity 51.0%; Pred. No. 1.2e-06;
Matches 123; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy    480 CAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGC 539
Db    560 CAAGCTTCCTGACAATGTCACCTTTGAGGAAGGCGCCCTGATCGAGCCACTTTCTGTGGG 619

Qy    540 CTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGG 599
Db    620 GATCCATGCCTGCAGGAGAGGCGGAGTTACCCTGGGACACAAGGTCCTTGTGTGTGGAGC 679

Qy    600 AGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGT 659
Db    680 TGGGCCAATCGGGATGGTCACTTTGCTCGTGGCCAAAGCAATGGGAGCAGCTCAAGTAGT 739

Qy    660 GTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTT 719
Db    740 GGTGACTGATCTGTCTGCTACCCGATTGTCCAAAGCCAAGGAGATTGGGGCTGATTTAGT 799

Qy    720 C 720
Db    800 C 800
```

RESULT 12
US-09-880-107-2189
; Sequence 2189, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2189
; LENGTH: 2519
; TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L29008
US-09-880-107-2189

Query Match 4.6%; Score 52.2; DB 10; Length 2519;
Best Local Similarity 51.0%; Pred. No. 1.2e-06;
Matches 123; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy    480 CAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGC 539
          |||||| || |||    |  ||   ||    || || ||| | |||||  |||||||
Db    560 CAAGCTTCCTGACAATGTCACCTTTGAGGAAGGCGCCCTGATCGAGCCACTTTCTGTGGG 619

Qy    540 CTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGG 599
            |||||        | |   || |    |  || || | || ||| ||   |||||
Db    620 GATCCATGCCTGCAGGAGAGGCGGAGTTACCCTGGGACACAAGGTCCTTGTGTGTGGAGC 679

Qy    600 AGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGT 659
           || || |||||  | | |||  | || |      |||       || |   || | ||
Db    680 TGGGCCAATCGGGATGGTCACTTTGCTCGTGGCCAAAGCAATGGGAGCAGCTCAAGTAGT 739

Qy    660 GTGTTCCGAGCCGGCCTTGATCAGAAGACAGTTTGCAAAGGAACTGGGCGCTGAAGTGTT 719
          |    | || | | |    | | ||      || |||||  | || |||||  |  |
Db    740 GGTGACTGATCTGTCTGCTACCCGATTGTCCAAAGCCAAGGAGATTGGGGCTGATTTAGT 799

Qy    720 C 720
          |
Db    800 C 800
```

RESULT 13
US-09-974-300-551
; Sequence 551, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 551
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-551

Query Match 4.0%; Score 45.8; DB 10; Length 1014;
Best Local Similarity 54.1%; Pred. No. 0.0001;
Matches 139; Conservative 0; Mismatches 112; Indels 6; Gaps 2;

```
Qy    358 GGGCTCTGTATGGCCTGCCAGAGCGGATCTCCGAACTGCTGTGCGTCGCTGAGCTTCTGC 417
          |||   || || |||||||  || ||||   || ||||||||  |     |||| | ||  |
Db    268 GGGGCATGCATCGCCTGCCGGAACGGACGCCCCAACTGCTGCGTCAATCTGAACGTCCTC. 327

Qy    418 GGTTTGGGTGGTGCCAGCGGCGGTTTTGCCGAGTACGTCGTTTACGGTGAGGACCACATG 477
          |||   |    | |   ||| ||  |    || ||| |      |   |  ||||||  ||
Db    328 GGT---GTCCATACAGACGGGGGAATGCGGGAATACATAAACGTCCCGGCAGACCACCTG 384

Qy    478 GTCAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTT 537
          | ||      ||  | ||  | |  |||    |  ||  | |||||   | |     |
Db    385 CTAAAAACAGAAGGATTGACGCTCAGCGA---AGCTGCTGTTGTTGAATGTTTAAGCATC 441

Qy    538 GCCTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGA 597
          | |   || |||| ||| | || |    || |    ||  | ||||  || ||| |||||
Db    442 GGCGCCCACGCTGCAGAAAGGGCAAATATCAAAAAAGGGGAAACGGTGCTTGTTGTTGGA 501
```

Qy 598 GGAGGTCCTATCGGCCT 614
| || || ||||| ||
Db 502 GCGGGGCCCATCGGGCT 518

RESULT 14
US-10-025-380-153
; Sequence 153, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-025-380-153

Query Match 3.9%; Score 45; DB 9; Length 513;
Best Local Similarity 51.2%; Pred. No. 0.00013;
Matches 105; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy    480 CAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGC 539
          |||||| || |||    |  ||   ||    || || ||| | |||||  |||||||
Db    304 CAAGCTTCCTGACAATGTCACCTTTGAGGAAGGCGCCCTGATCGAGCCACTTTCTGTGGG 363

Qy    540 CTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGG 599
            |||||        | |   || |    |  || || | || ||| ||   |||||
Db    364 GATCCATGCCTGCAGGAGAGGCGGAGTTACCCTGGGACACAAGGTCCTTGTGTGTGGAGC 423

Qy    600 AGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGT 659
           || || |||||  | | |||  | || |      |||       || |   || | ||
Db    424 TGGGCCAATCGGGATGGTCACTTTGCTCGTGGCCAAAGCAATGGGAGCAGCTCAAGTAGT 483

Qy    660 GTGTTCCGAGCCGGCCTTGATCAGA 684
          |    | || | | |    | | ||
Db    484 GGTGACTGATCTGTCTGCTACCCGA 508
```

RESULT 15
US-09-922-217-153
; Sequence 153, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong

```
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153
;   LENGTH: 513
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-217-153

  Query Match             3.9%;  Score 45;  DB 10;  Length 513;
  Best Local Similarity   51.2%;  Pred. No. 0.00013;
  Matches  105;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy    480 CAAGCTGCCAGACTCGATTCCCGACGATATTGGAGCACTGGTTGAGCCTATTTCTGTTGC 539
          |||||| || |||    |  ||   ||    || || ||| | |||||  |||||||
Db    304 CAAGCTTCCTGACAATGTCACCTTTGAGGAAGGCGCCCTGATCGAGCCACTTTCTGTGGG 363

Qy    540 CTGGCATGCTGTTGAACGCGCTAGATTCCAGCCTGGTCAGACGGCCCTGGTTCTTGGAGG 599
            |||||        | |   || |    |  || || | || ||| ||   |||||
Db    364 GATCCATGCCTGCAGGAGAGGCGGAGTTACCCTGGGACACAAGGTCCTTGTGTGTGGAGC 423

Qy    600 AGGTCCTATCGGCCTTGCCACCATTCTTGCTCTGCAAGGCCATCATGCGGGCAAAATTGT 659
          || || |||||  | | |||  | || |      |||        || |   || | ||
Db    424 TGGGCCAATCGGGATGGTCACTTTGCTCGTGGCCAAAGCAATGGGAGCAGCTCAAGTAGT 483

Qy    660 GTGTTCCGAGCCGGCCTTGATCAGA 684
          |    | || | | |    | | ||
Db    484 GGTGACTGATCTGTCTGCTACCCGA 508


Search completed: May  2, 2003, 12:08:43
Job time : 177 secs
```